Compressive Optical Deflectometric Tomography: A Constrained Total-Variation Minimization Approach.

Adriana González[1]

ICTEAM/IMCN, Université catholique de Louvain. Louvain-la-Neuve, Belgium.

Laurent Jacques

ICTEAM, Université catholique de Louvain. Louvain-la-Neuve, Belgium.

Christophe De Vleeschouwer

ICTEAM, Université catholique de Louvain. Louvain-la-Neuve, Belgium.

Philippe Antoine

IMCN, Université catholique de Louvain. Louvain-la-Neuve, Belgium.

Lambda-X SA, Rue de l'Industrie 37. Nivelles, Belgium.

September 5, 2012

Abstract

Optical Deflectometric Tomography (ODT) provides an accurate characterization of transparent materials whose complex surfaces present a real challenge for manufacture and control. In ODT, the refractive index map (RIM) of a transparent object is reconstructed by measuring light deflection under multiple orientations. We show that this imaging modality can be made *compressive*, *i.e.*, a correct RIM reconstruction is achievable with far less observations than required by traditional Filtered Back Projection (FBP) methods. Assuming a cartoon-shape RIM model, this reconstruction is driven by minimizing the map Total-Variation under a fidelity constraint with the available observations. Moreover, two other realistic assumptions are added to improve the stability of our approach: the map positivity and a frontier condition. Numerically, our method relies on an accurate ODT sensing model and on a primal-dual minimization scheme, including easily the sensing operator and the proposed RIM constraints. We conclude this paper by demonstrating the power of our method on synthetic and experimental data under various compressive scenarios. In particular, the compressiveness of the stabilized ODT problem is demonstrated by observing a typical gain of 20 dB compared to FBP at only 5% of 360 incident light angles for moderately noisy sensing.

Keywords: Optical Deflectometric Tomography, Refractive index map, Total Variation Minimization, NFFT, Primal-Dual Optimization.

1   INTRODUCTION

Optical Deflectometric Tomography (ODT) is an imaging modality that aims at reconstructing the spatial distribution of the refractive index of a transparent object from the deviation of the light passing through the object. By reconstructing the refractive index map (RIM) we are

---

[1] {adriana.gonzalez, laurent.jacques, christophe.devleeschouwer, ph.antoine}@uclouvain.be

able to optically characterize transparent materials like optical fibers or multifocal intra-ocular lenses (IOL), which is of great importance in manufacture and control processes.

ODT is attractive for its high sensitivity and effective detection of local details and object contours. Also, ODT is insensitive to vibrations and it does not require coherent light. Moreover, compared to interferometry, ODT can be applied to objects with higher refractive index difference with respect to the surrounding solution.

The deflectometer used for the data acquisition is based on the phase-shifting Schlieren method [17]. For each orientation of the sample, two-dimensional (2-D) mappings of local light deviations are measured. As these light deviations are proportional to the transverse gradient of the RIM integrated along the light ray, ODT is able to reconstruct the RIM from the angle measurements.

First works on deflectometric tomography [3, 16] have focused on the use of common reconstruction techniques like the Filtered Back Projection (FBP). They have proved that FBP provides an accurate estimation of the RIM when sufficient amount of object orientations is available. However, when we consider a scenario with a limited number of light incident angles, and in the presence of noise in ODT observations, FBP induces the apparition of spurious artifacts in the estimated RIM, lowering the reconstruction quality.

Inspired by the Compressed Sensing (CS) paradigm [6, 12], which demonstrates that few measurements are enough for an accurate reconstruction of sparse signals, recent works in ODT have started to exploit sparsity to reconstruct the RIM from few acquisitions, *i.e.*, in the presence of an ill-posed inverse problem. Foumouo et al. [17] and Antoine et al. [1] have used a sparse representation in a B-splines basis and regularized the problem using the ℓ_1-norm. The reconstruction was performed using iterative schemes and the results show that, although the method is capable of reproducing the shape of spatially localized objets, the image dynamics is not well preserved and the borders are smoothened.

Although sparsity based methods are new in ODT, they have been used in other applications, such as Magnetic Resonance Imaging (MRI) [25], Absorption Tomography (AT) [30, 32], Radio Interferometry [35] and Phase-Contrast Tomography [10, 11], for the image reconstruction when few amount of acquisitions is available. More details are given in Sec. 4.

An additional problem that rises in all physical applications is the estimation of the sensing operator that fits better the physical acquisition. Most operators present a non ideal behavior, which conditions the numerical methods to solve the inverse problem. In tomographic problems, this operator requires to map spatial data in a Cartesian grid to Fourier data in a Polar grid. Previous works have used gridding techniques to interpolate data from a polar to a cartesian or pseudo polar grid before applying the Fourier Transform [19, 36]. However, the error introduced when using these techniques is not bounded and introduces an uncontrolled distortion.

1.1 Contribution

In this work, we show that ODT can be compressive and robust to Gaussian noise. We propose a constrained method based on the minimization of the Total-Variation (TV) norm that provides high quality reconstruction results even when few acquisitions are available and in the presence of high level of Gaussian noise.

We assume the RIM is composed by slowly varying areas separated by sharp transitions corresponding to material interfaces. Such a behavior is known to be represented by spatial functions having a small TV norm. Other sparsity basis, such as the B-splines basis in [1, 17], tend to be complicated to use and do not provide a complete characterization of the RIM, while the TV norm represents a more simple and accurate model of the actual RIM.

Therefore, we propose the use of the TV norm to regularize the ill-posed inverse problem of

estimating the RIM from low amount of noisy data. To account for the noise and the raw data consistency, we add an ℓ_2 data fidelity term adapted to Gaussian and uniformly distributed noise.

The proposed method offers the flexibility to work with more than one constraint. In this work, we add two more constraints based on some general prior information on the RIM in order to converge to an optimal solution: (i) the RIM is positive everywhere and (ii) the object is completely contained in the field of view of the ODT experiment. These added constraints are proved to provide an unique solution to our ODT problem.

As for the operator, we use the NFFT algorithm[2], a fast computation of the non-equispaced DFT. This algorithm provides an efficient estimation of the polar Fourier transform with a controlled distortion regarding the true polar transform.

The compressive ODT problem is solved by means of the primal-dual algorithm proposed by Chambolle et al. [8]. The results are compared with a common FBP approach, showing the proposed method outperforms FBP in terms of compressiveness, noise robustness and reconstruction quality.

1.2 Outline

In Sec. 2 we provide a brief background on optical deflectometric tomography, describing also the experimental setup used for the data acquisition. Then, the ODT discrete model is presented in Sec. 3. In Sec. 4 we depict related works on tomographic reconstruction, which provide a basis on the methods adopted to recover the RIM: the commonly used FBP method and a proposed regularized method that we call TV-ℓ_2. These methods are described in Sec. 5. In Sec. 6 we present the identification and estimation of the noise in both synthetic and experimental data, and the analysis of the noise impact when comparing common absorption tomography and deflection tomography. Sec. 7 presents the numerical methods we use to recover the RIM from the noisy measurements by means of the proposed regularized formulation. Finally, in Sec. 8 some reconstruction results are shown, focusing first on the comparison between common tomographic and ODT reconstructions, and then on the comparison of the reconstruction methods on the basis of quality reconstruction and convergence for both synthetic and experimental data.

1.3 Conventions

Most of domain dimensions, *e.g.*, $M, N, N_1, N_2, \ldots$, are denoted by capital roman letters. Vectors and matrices are associated to bold symbols while lowercase light letters are associated to scalar values. To give an example, for some dimensions $D, D' \in \mathbb{N}_0$, we write $\boldsymbol{u} = (u_1, \cdots, u_D)^T \in \mathbb{R}^D$ (using the transposition $(\cdot)^T$) or we define the matrix $\boldsymbol{\Phi} \in \mathbb{R}^{D' \times D}$ with matrix entry $\Phi_{ij} \in \mathbb{R}$. Therefore, the ith component of a vector $\boldsymbol{u}$ is denoted either u_i or $(\boldsymbol{u})_i$, the later notation being useful if $\boldsymbol{u}$ brings some extra notations (*e.g.*, $\boldsymbol{u}_S$), while $\boldsymbol{u}_i$ will denote for instance the ith vector of a set of vectors $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots\}$. The identity matrix in $\mathbb{R}^D$ is denoted $\mathbb{I}^D$, or simply $\mathbb{I}$ when its dimension is clear from the context. The set of indices in $\mathbb{R}^D$ is $[D] = \{1, \cdots, D\}$. Scalar product between two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^D$ for some dimension $D \in \mathbb{N}$ is denoted equivalently by $\boldsymbol{u}^T \boldsymbol{v} = \boldsymbol{u} \cdot \boldsymbol{v} = \langle \boldsymbol{u}, \boldsymbol{v} \rangle$. For any $p \geqslant 1$, $\| \cdot \|_p$ represents the ℓ_p-norm such that $\|\boldsymbol{u}\|_p^p = \sum_i |u_i|^p$ with $\| \cdot \| = \| \cdot \|_2$. The notation $\|\boldsymbol{u}\|_0 = \# \operatorname{supp} \boldsymbol{u}$ represents the cardinality of the support $\operatorname{supp} \boldsymbol{u} = \{i : u_i \neq 0\} \subset [D]$. For a subset $\mathcal{S} \subset [D]$, given $\boldsymbol{u} \in \mathbb{R}^D$ and $\boldsymbol{\Phi} \in \mathbb{R}^{D' \times D}$, $\boldsymbol{u}_\mathcal{S} \in \mathbb{R}^{\#\mathcal{S}}$ (or $\boldsymbol{\Phi}_\mathcal{S}$) denotes the vector (resp. the matrix) obtained by retaining the components (resp. columns) of $\boldsymbol{u}$ (resp. $\boldsymbol{\Phi}$) belonging to $\mathcal{S}$. Alternatively, we have $\boldsymbol{u}_\mathcal{S} = \boldsymbol{R}_\mathcal{S} \boldsymbol{u}$

---

[2]This toolbox is freely available here `http://www-user.tu-chemnitz.de/~potts/nfft/`
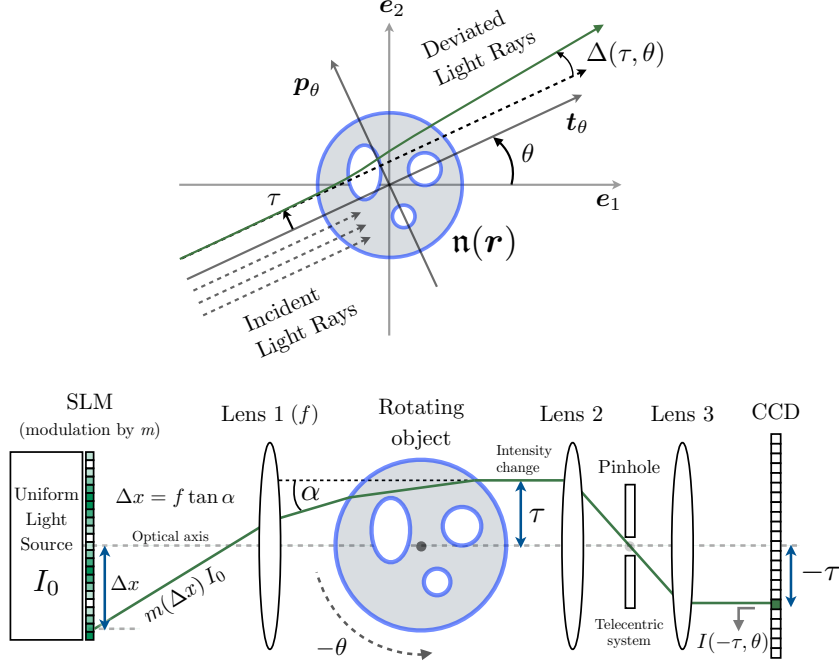
Figure 1: (top) The Deflectometric model. (bottom) A scheme of a Phase-Shifting Schlieren Deflectometer (in refraction mode) measuring light ray deflection angles by encoding light deviation into light intensity variations.

or $\boldsymbol{\Phi}_{\mathcal{S}} = \boldsymbol{\Phi} \boldsymbol{R}_{\mathcal{S}}^T$ where $\boldsymbol{R}_{\mathcal{S}} := (\mathbb{I}_{\mathcal{S}})^T \in \{0, 1\}^{\#\mathcal{S} \times D}$ is the *restriction* operator. The kernel of a matrix $\boldsymbol{\Phi}$ is defined as $\ker(\boldsymbol{\Phi}) := \{\boldsymbol{x} \in \mathbb{R}^D : \boldsymbol{\Phi}\boldsymbol{x} = 0, \boldsymbol{x} \neq 0\}$. The norm of an operator $\boldsymbol{K}$ is defined as $\|\boldsymbol{K}\| = \max\{\|\boldsymbol{K}\boldsymbol{x}\| : \boldsymbol{x} \in \mathbb{R}^N \text{ with } \|\boldsymbol{x}\| = 1\}$. We denote $\Gamma^0(\mathcal{V})$ the class of proper, convex and lower-semicontinuous functions from a finite dimensional vector space $\mathcal{V}$ (*e.g.*, $\mathbb{R}^D$) to $(-\infty, +\infty]$ [9, 15]. We denote by $(\boldsymbol{x})_+$ the non-negativity thresholding function, which is defined componentwise as $(x_i)_+ = (x_i + |x_i|)/2$. We denote by $\iota_{\mathcal{C}}(\boldsymbol{x})$ the indicator function of the set $\mathcal{C}$, which is equal to 0 if $\boldsymbol{x} \in \mathcal{C}$ and $+\infty$ otherwise. We denote by $\text{int}\,\Omega$ the interior of a set Ω, which consists of all points of Ω that do not belong to the boundary $\partial\Omega$ of Ω.

2 OPTICAL DEFLECTOMETRIC TOMOGRAPHY

In this section, the principles of optical deflectometric tomography are explained and completed with a brief description of the experimental setup used for actual deflectometric acquisition.

2.1 Principles

Optical Deflectometric Tomography aims at infering the refractive index spatial distribution (or refractive index map – RIM) of a transparent object. This is achieved by measuring, under various incident angles, the deflection angles of multiple parallel light rays when passing through this transparent object (see Fig. 1-(top)). The (indirect) observation of the RIM, allowing its further reconstruction, is guaranteed by the relation between the total light ray deflection angle and the integration of the transverse gradient of the RIM along the light ray path [3].

In this work, we restrict ourselves to two-dimensional ODT by assuming that the refractive index $\mathfrak{n}$ of the observed object is constant along the $\boldsymbol{e}_3$-axis for a convenient coordinate system $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$, *i.e.*, $\partial\mathfrak{n}(\boldsymbol{r})/\partial r_3 = 0$ (with $\boldsymbol{r} = (r_1, r_2, r_3)^T \in \mathbb{R}^3$) and deflections occur in the $\boldsymbol{e}_1$-$\boldsymbol{e}_2$ plane. This strong restriction can be softened by assuming that the variations of $\mathfrak{n}$ along the

4

e_3 axis are slow compared to those occurring in the $\{e_1, e_2\}$ domain.

Consequently, the direction e_3 can be forgotten. Given the refractive index $\mathfrak{n} : r = (r_1, r_2)^T \in \mathbb{R}^2 \to \mathfrak{n}(r)$, for a particular light ray trajectory $\mathcal{R} = \{r(s) : s \in \mathbb{R}\} \subset \mathbb{R}^2$ parametrized by $r(s) = (r_1(s), r_2(s))^T \in \mathbb{R}^2$ with s describing its *curvilinear* parameter[3], the deflection angle is provided by [22]

$$\Delta(\mathcal{R}; \mathfrak{n}) \;=\; \tfrac{1}{\mathfrak{n}_r} \int_{\mathcal{R}} \left[\boldsymbol{\nabla}\mathfrak{n}\big(r(s)\big) \cdot \boldsymbol{p}(s) \right] \mathrm{d}s, \tag{1}$$

where $\mathfrak{n}_r$ is the (constant) reference refractive index of the surrounding medium, $\boldsymbol{p}(s)$ is the unit normal vector to the curve $\mathcal{R}$ in $r(s)$ such that $\boldsymbol{p}(s) \cdot \frac{\mathrm{d}}{\mathrm{d}s} r(s) = 0$, and the quantity $\boldsymbol{\nabla}\mathfrak{n}\big(r(s)\big) \cdot \boldsymbol{p}(s)$ is actually the transverse gradient of $\mathfrak{n}$ along $\mathcal{R}$ (the directional derivative of $\mathfrak{n}$ along $\boldsymbol{p}$).

Of course the light ray trajectory $\mathcal{R}$ depends itself on $\mathfrak{n}$ as computed by the *light ray equation* $\frac{\mathrm{d}}{\mathrm{d}s}\big(\mathfrak{n}\frac{\mathrm{d}}{\mathrm{d}s}r(s)\big) = \boldsymbol{\nabla}\mathfrak{n}$ established from the eikonal equation [4]. However, for small deflection angles Δ, we can adopt the (first order) *paraxial* approximation and assume the trajectory $\mathcal{R}$ is a straight line.

In this simplified model, the 2-D RIM is measured by the deflection angle $\Delta(\tau, \theta; \mathfrak{n})$ of a light ray $\mathcal{R}(\theta, \tau) = \{r \in \mathbb{R}^2 : r \cdot \boldsymbol{p}_\theta = \tau\}$, where $\tau \in \mathbb{R}$ is the *affine* parameter determining the distance between $\mathcal{R}$ and the origin, $\theta \in [0, 2\pi)$ is the incident angle of $\mathcal{R}$ with the e_1 axis, and $\boldsymbol{p}_\theta = (-\sin\theta, \cos\theta)^T$ is the (constant) perpendicular vector to the light ray direction $\boldsymbol{t}_\theta = (\cos\theta, \sin\theta)^T$.

Consequently, removing the explicit trajectory dependence in Δ, the simplified ODT model depicted in Fig. 1-(top) reads

$$\begin{aligned}
\Delta(\tau, \theta) &:= \tfrac{1}{\mathfrak{n}_r} \int_{\mathbb{R}} \boldsymbol{\nabla}\mathfrak{n}(r_{\tau,\theta}(s)) \cdot \boldsymbol{p}_\theta \, \mathrm{d}s \\
&= \tfrac{1}{\mathfrak{n}_r} \int_{\mathbb{R}^2} \big(\boldsymbol{\nabla}\mathfrak{n}(r) \cdot \boldsymbol{p}_\theta\big) \, \delta(\tau - r \cdot \boldsymbol{p}_\theta) \, \mathrm{d}^2 r,
\end{aligned} \tag{2}$$

where $r_{\tau,\theta}(s) = s\boldsymbol{t}_\theta + \tau\boldsymbol{p}_\theta \in \mathcal{R}$, and the Dirac distribution turns the line integral into an integration over $\mathbb{R}^2$. In short, the above equation relates the deflection angle $\Delta(\tau, \theta)$ to the Radon transform of the transverse gradient of $\mathfrak{n}$ within the paraxial approximation.

As for traditional parallel absorption tomography [23], there exists a Deflectometric Fourier Slice Theorem (DFST) that relates the 1-D (radial) Fourier transform of the deflection angle along the *affine parameter* τ, i.e.,

$$y(\omega, \theta) \;:=\; \int_{\mathbb{R}} \Delta(\tau, \theta) \, e^{-2\pi i \tau \omega} \mathrm{d}\tau, \tag{3}$$

to the 2-D Fourier transform of the RIM. Mathematically, the DFST establishes the following equivalence [19] (proved in Appendix A):

$$y(\omega, \theta) \;=\; \tfrac{2\pi i \omega}{\mathfrak{n}_r} \widehat{\mathfrak{n}}\big(\omega \, \boldsymbol{p}_\theta\big), \tag{4}$$

where $\hat{\mathfrak{n}}(\boldsymbol{k}) = \int_{\mathbb{R}^2} \mathfrak{n}(r) \, e^{-2\pi i \boldsymbol{k} \cdot r} \, \mathrm{d}^2 r$ stands for the 2-D Fourier transform of $\mathfrak{n}$. Hereafter, the value ω is called *affine frequency*.

We may remark from (4) that there are different ways to cover the 2-D frequency plane with deflectometric measurements. This can be observed from the following symmetry relations for

---

[3]By a slight abuse of notation, r denotes any points in $\mathbb{R}^2$ while $r(s)$ represents a particular curve in $\mathbb{R}^2$ parametrized by $s \in \mathbb{R}$.

$\mathfrak{n} \in \mathbb{R}$, $\tau, \omega \in \mathbb{R}$, $\theta \in [0, 2\pi)$:

$$\Delta(\tau, (\theta + \pi) \mod 2\pi) = -\Delta(-\tau, \theta), \tag{5a}$$

$$y(\omega, (\theta + \pi) \mod 2\pi) = -y(-\omega, \theta), \tag{5b}$$

$$y(\omega, \theta) = y^*(-\omega, \theta), \tag{5c}$$

where the two first relations come from the change $\boldsymbol{p}_{\theta+\pi} = -\boldsymbol{p}_\theta$ in (2) and (4), and the last equation is the well known Fourier conjugate symmetry for real spatial functions.

In particular, from the symmetry (5c), we can restrict ω to positive values by taking θ in the whole circle $[0, 2\pi)$. Or alternatively, from (5b) and (5c), we can deduce $y(\omega, \theta) = -y^*(\omega, (\theta + \pi) \mod 2\pi)$ and restrict θ to the half circle $[0, \pi)$ with $\omega \in \mathbb{R}$. We insist on the fact that this symmetry is only preserved when the paraxial approximation is validated.

When comparing the relation (4) with the standard tomographic Fourier Slice theorem (FST) [23], the main difference is provided by the transverse gradient in the deflectometric relation (2), which results in multiplying by $2\pi i \omega / \mathfrak{n}_r$ the RIM Fourier transform. In particular, from (2) or from (4) (since ω vanishes on the frequency origin) we see that the ODT sensing is blind to constant RIM. As we will see in Sec. 5, this has an impact on the formulation of the RIM reconstruction since the RIM can only be estimated relatively to the reference RIM $\mathfrak{n}_r$.

Let us conclude this section by insisting on the impact of the equivalence (4). Similarly to the use of the Fourier Slice theorem in common (absorption) tomography, (4) is of great importance for defining a discrete ODT sensing model which can be computed efficiently in the Fourier domain given a discretized refractive index map $\mathfrak{n}$.

2.2 Deflection Measurements

Experimentally, the deflection angles Δ can be measured by phase-shifting Schlieren deflection tomography (schematically represented in Fig. 1-(bottom)). We briefly explain this system for the sake of completeness in order to set the experimental background surrounding the actual deflection measurement process. More details can be found in [3, 17, 19, 21].

This system proceeds by encoding light ray deflection α in intensity variations. A transparent object is illuminated with an incoherent uniform light source I_0 modulated by a sinusoidal pattern m using a Spatial Light Modulator (SLM). From classical optics, the light deviation angle α is related to a phase shift $\Delta x = f \tan \alpha$, where f is the focal length of Lens 1. This phase shift is associated with the intensity variation thanks to the modulation m as $I(-\tau, \theta) = m(\Delta x)I_0$. These intensity variations are processed by phase shifting methods for recovering the deflection measurements $\alpha = \Delta(\tau, \theta)$ for each couple of parameters $(\tau, \theta) \in \mathbb{R} \times [0, 2\pi]$. Up to some linear coordinate rescaling, the *affine* parameter τ corresponds to the horizontal pixel coordinate in the 2-D CCD detector collecting light (assuming the object refractive index constant along the CCD vertical direction). This correspondence is implicitly allowed by the telecentric system formed by the combination of Lens 2 and 3. The pinhole guarantees that only parallel light rays outgoing from the object are collected. Rather than rotating the whole incident light beam around the object, it is this one which is rotated by an angle $-\theta$ along an axis parallel to the CCD pixel vertical direction [3]. Finally, since the system is invariant under time inversion, *i.e.*, under light progression inversion, measuring the deflection angle α in Fig. 1-(bottom) is equivalent to measuring the same angle in Fig. 1-(top).

3 DISCRETE FORWARD MODEL

In order to reconstruct efficiently the RIM from ODT measurements, recorded data must be treated appropriately considering, jointly, the data discretization, the polar geometry of the

ODT sensing and unavoidable measurement noises. In this section, the discrete formulation of the ODT sensing and the construction of the forward model from the recorded data are presented.

3.1 Discrete domains

Let us first assume that the object of interest is fully contained in a square field-of-view (or FoV) $\Omega \subset \mathbb{R}^2$ centered on the spatial origin. The physical dimensions of this FoV can be provided by the Deflectometric device itself. In other words, the RIM is constant and equal to the reference index $\mathfrak{n}_r$ outside of Ω. This involves also that the deflection measurement vanishes, i.e., $\Delta(\tau, \omega) = 0$, if $|\tau|$ is bigger than the typical width of Ω in a section of direction $\theta + \pi/2$.

We can consider a spatial sampling of Ω as follows. We define a $N_0 \times N_0$ 2-D Cartesian grid of $N := N_0^2$ pixels as

$$\mathcal{C}_N = \{\boldsymbol{r}_{m,n} := (m\,\delta r, n\,\delta r) : -N_0/2 \leqslant m, n < N_0/2\},$$

where the spatial spacing δr is adjusted to Ω and to the resolution by imposing $\Omega = [-\frac{1}{2}N_0\,\delta r, \frac{1}{2}N_0\,\delta r] \times [-\frac{1}{2}N_0\,\delta r, \frac{1}{2}N_0\,\delta r]$.

Second, as the deflectometric experiments provide evenly sampled variables τ and θ, Δ is measured on a (signed) regular polar coordinate grid

$$\mathcal{P}_M := \{(\tau_s, \theta_t) : -(N_\tau/2) \leqslant s < (N_\tau/2), 0 \leqslant t < N_\theta\}, \ \tau_s := s\,\delta\tau, \ \theta_t := t\,\delta\theta,$$

of size $M := N_\tau N_\theta$, with N_τ the number of parallel light rays passing through the object (assumed even), N_θ the number of incident angles in ODT sensing, $\delta\tau$ and $\delta\theta = \pi/N_\theta$ the distance between two consecutive affine parameters and angles respectively[4].

The value $\delta\tau$ can be known experimentally from the pixel size of the CCD detector in a Schlieren Deflectometer (Sec. 2.2). Moreover, the value $\delta\tau$ and the resolution N_τ are also related to the FoV Ω so that $\delta\tau N_\tau \approx \delta r N_0$. Since there is no reason to ask more resolution to the sampling $\mathcal{C}_N$ than the available in the affine variations of the ODT measurements, we will work with $\delta\tau \approx \delta r$ and $N_\tau \approx N_0$.

Third, in this discretized setting, the affine frequency ω in (4) must also be sampled with N_τ values. As described in the next section, this comes from the replacement of the (radial) Fourier transform in (3) by its discrete counterpart. This leads to a (signed) frequency polar grid of same size

$$\hat{\mathcal{P}}_M := \{(\omega_{s'}, \theta_t) : -(N_\tau/2) \leqslant s' < (N_\tau/2), 0 \leqslant t < N_\theta\}, \ \omega_{s'} := s'\,\delta\omega, \ \theta_t := t\,\delta\theta,$$

with $\delta\omega = 1/(N_\tau\delta\tau)$. As it will become clearer in the following, only half of this polar grid will be necessary to bring independent observations of the RIM, i.e., we will often work on

$$\hat{\mathcal{P}}_M^+ := \{(\omega_{s'}, \theta_t) : 0 \leqslant s' < (N_\tau/2), 0 \leqslant t < N_\theta\},$$

with $\#\hat{\mathcal{P}}_M^+ = M/2$.

3.2 Discretized Functions

From the discrete domains defined above, the continuous RIM $\mathfrak{n}$ observed in the experimental FoV is discretized into a set of N values $\mathfrak{n}(\boldsymbol{r}_{m,n})$ from the coordinates $\boldsymbol{r}_{m,n} \in \mathcal{C}_N$. This description can always be arranged into a one dimensional N-length vector $\mathfrak{n} \in \mathbb{R}^N$, given a convenient

---

[4]Notice that $\delta\theta$ is not set to $2\pi/N_\theta$ since τ is allowed to be negative.

mapping between the components indices of $\mathfrak{n} = (\mathfrak{n}_1, \cdots, \mathfrak{n}_N)^T$ and the pixel coordinates in $\mathcal{C}_N$. This "vectorization" provides us a simplified representation of any linear operator acting on the sampled RIM $\mathfrak{n}$ as a matrix multiplying $\mathfrak{n}$.

For the different functions discretized on $\mathcal{P}_M$ or on $\hat{\mathcal{P}}_M$, we use the same vectorization trick, namely, a function u defined on $\mathcal{C}_N$ and sampled on $\mathcal{P}_M$ is associated to a vector $\boldsymbol{u} \in \mathbb{R}^M$ with the right correspondence between the components of $\boldsymbol{u}$ and the polar coordinates in $\mathcal{P}_M$ (and similarly for function defined in the Fourier domain and sampled on $\hat{\mathcal{P}}_M$).

Therefore, the ODT observations $\{\Delta(\tau, \theta) : (\tau, \theta) \in \mathcal{P}_M\}$ are gathered in a vector $\boldsymbol{z} \in \mathbb{R}^M$ with $z_j = \Delta(\tau_s, \theta_t)$ for a certain index mapping $j = j(s, t) \in [M]$. In this discrete representation, the equivalent transformation of (3) reads

$$\boldsymbol{y}_{\text{comp}} = (\sqrt{N_\tau}\,\delta\tau)\,\boldsymbol{F}^{\text{rad}}\boldsymbol{z}, \tag{6}$$

where $\boldsymbol{y}_{\text{comp}} \in \mathbb{C}^M$ is associated to a (vectorized) sampling of y on $\hat{\mathcal{P}}_M$, and $\boldsymbol{F}^{\text{rad}} : \mathbb{R}^M \to \mathbb{C}^M$ performs a 1-D DFT on the radial τ-variations of $\boldsymbol{z}$, $i.e.$,

$$(\boldsymbol{F}^{\text{rad}}\boldsymbol{z})_{k(s',t)} = \tfrac{1}{\sqrt{N_\tau}} \sum_s z_{j(s,t)}\, e^{-2\pi i s s'/N_\tau},$$

for a vectorized index $k = k(s', t)$. In other words, if $\delta\tau$ is sufficiently small, $e.g.$, if $\Delta(\tau, \theta)$ is band-limited with a cut-off frequency smaller than $1/\delta\tau = N_\tau \delta\omega$, then

$$y_{k(s',t)} = \tfrac{\sqrt{N_\tau}\,\delta\tau}{\sqrt{N_\tau}} \sum_s \Delta(\tau_s, \theta_t)\, e^{-2\pi i (s\delta\tau)(s'/\delta\tau N_\tau)}$$

$$= \sum_s \Delta(\tau_s, \theta_t)\, e^{-2\pi i \tau_s \omega_{s'}}\,\delta\tau$$

$$\approx\ y(\omega_{s'}, \theta_t),$$

using a Riemann sum approximation of (3) and knowing that Δ vanishes outside of Ω.

Despite the fact that $\boldsymbol{y}_{\text{comp}}$ belongs $\mathbb{C}^M \simeq \mathbb{R}^{2M}$, this vector brings only M independent real observations of $\mathfrak{n} \in \mathbb{R}^N$. This is due to the central symmetry (5c) induced by the realness of $\boldsymbol{z}$ and which allows us to consider $\boldsymbol{y}_{\text{comp}}$ only on $\hat{\mathcal{P}}_M^+$.

This is clarified by the definition of the useful operator $\boldsymbol{\Theta} : \mathbb{C}^M \to \mathbb{R}^M$ which perform the two following linear operations. First, it restricts any vector $\boldsymbol{\xi} \in \mathbb{C}^M$ to the indices associated to the half grid $\hat{\mathcal{P}}_M^+$. Second, it appends the $M/2$ imaginary values of the restricted vector to its $M/2$ real values in order to form a M-length vector in $\mathbb{R}^M$. The adjoint operation $\boldsymbol{\Theta}^*$, which is also the inverse of $\boldsymbol{\Theta}$ for vectors in $\mathbb{C}^M$ respecting the Hermitian symmetry, is obtained easily by first reforming a $M/2$-length complex vector from the separated real and imaginary parts, and by inserting the results in $\mathbb{C}^M$ according to the indices of $\hat{\mathcal{P}}_M^+$ and by completing the information in $\hat{\mathcal{P}}_M \setminus \hat{\mathcal{P}}_M^+$ with the central symmetry (5c).

Consequently, thanks to $\boldsymbol{\Theta}$, we can form the real vector

$$\boldsymbol{y} = \boldsymbol{\Theta}\,\boldsymbol{y}_{\text{comp}} = (\sqrt{N_\tau}\,\delta\tau)\,(\boldsymbol{\Theta}\,\boldsymbol{F}^{\text{rad}})\,\boldsymbol{z} \quad \in \mathbb{R}^M, \tag{7}$$

with $(\boldsymbol{\Theta}\,\boldsymbol{F}^{\text{rad}}) : \mathbb{R}^M \to \mathbb{R}^M$. We call $\boldsymbol{y}$ the Frequency Deflectometric Measurements (FDM) vector and we will most often use it as our direct source of ODT observations instead of $\boldsymbol{z}$.

3.3 Forward Model

We can now explain how we use the DFST relation (4) for defining the forward model that links any discrete 2-D RIM representation to its FDM vector.

In a previous work [19], the data available in the frequency polar grid $\hat{\mathcal{P}}_M$ were first interpolated to a Cartesian frequency grid in order to reconstruct the 2-D RIM using a Discrete Fourier Transform (through the FFT algorithm). However, the polar to Cartesian frequency interpolation introduced a hardly controlled colored distortion.

We prefer here the use of another operator $\boldsymbol{F} : \mathbb{R}^N \to \mathbb{C}^M$ performing a non-equispaced Discrete Fourier Transform (NDFT) for directly relating functions sampled on the Cartesian spatial grid $\mathcal{C}_N$ to those sampled on the polar frequency grid $\hat{\mathcal{P}}_M$.

More precisely, given a function $f : \Omega \to \mathbb{C}$ sampled on $\mathcal{C}_N$, the NDFT[5] computes

$$\hat{f}(\boldsymbol{k}) = \sum_{m=0}^{N_0-1} \sum_{n=0}^{N_0-1} f(\boldsymbol{r}_{m,n}) e^{-2\pi i \, \boldsymbol{k} \cdot \boldsymbol{r}_{m,n}}, \tag{8}$$

on the M nodes $\boldsymbol{k}$ of $\hat{\mathcal{P}}_M$. Gathering all the values of $\hat{f}$ and f into vectors $\hat{\boldsymbol{f}} \in \mathbb{C}^M$ and $\boldsymbol{f} \in \mathbb{R}^M$ respectively, this relation is conveniently summarized in the following equation

$$\hat{\boldsymbol{f}} = \boldsymbol{F}\boldsymbol{f}, \quad \text{with } F_{ij} = e^{-2\pi i \, \boldsymbol{k}_i \cdot \boldsymbol{r}_j}. \tag{9}$$

where the matrix $\boldsymbol{F} \in \mathbb{C}^{M \times N}$ stands for the linear NDFT operation. Its internal entry indexing follows the one of the components of $\boldsymbol{f}$ and $\hat{\boldsymbol{f}}$. We explain in Appendix B how the Non-equispaced Fast Fourier Transform (NFFT) algorithm allows us to compute efficiently in $\mathcal{O}(N \log(N/\epsilon))$ the multiplications $\boldsymbol{F}\boldsymbol{u}$ and $\boldsymbol{F}^*\boldsymbol{v}$ for any $\boldsymbol{u} \in \mathbb{R}^N$ and $\boldsymbol{v} \in \mathbb{C}^M$, with a controlled distortion ϵ with respect to the true NDFT.

The action of $\boldsymbol{F}$ on a discretized RIM $\mathfrak{n}$ is related to the continuous Fourier transform of $\mathfrak{n}$ as follows. Let $j = j(s',t) \in [M]$ be the jth point $\boldsymbol{k}_j$ of the (vectorized) grid $\hat{\mathcal{P}}_M$ associated to the polar coordinates $(\omega_{s'}, \theta_t)$. Then, for a sufficiently small δr,

$$(\boldsymbol{F}\mathfrak{n})_{j(s',t)} = \sum_{m=0}^{N_0-1} \sum_{n=0}^{N_0-1} \mathfrak{n}(\boldsymbol{r}_{m,n}) e^{-2\pi i \, \boldsymbol{k}_j \cdot \boldsymbol{r}_{m,n}} \approx \frac{1}{(\delta r)^2} \hat{\mathfrak{n}}(\omega_{s'} \boldsymbol{p}_{\theta_t}). \tag{10}$$

To take into account the multiplication by $2\pi i \, \omega / \mathfrak{n}_{\mathrm{r}}$ in (4) and the existence of a factor $1/(\delta r)^2$ in the equivalence (10), we introduce the diagonal operator $\boldsymbol{D} \in \mathbb{R}^{M \times M}$ defined as

$$\boldsymbol{D} = \frac{2\pi \, i (\delta r)^2}{\mathfrak{n}_{\mathrm{r}}} \operatorname{diag}(\omega_{(1)}, \cdots, \omega_{(M)}),$$

where $\omega_{(j)}$ refers to the ω-coordinate of the jth point of $\hat{\mathcal{P}}_M$, *i.e.*, if as before $j = j(s',t)$ is associated to $(\omega_{s'}, \theta_t) \in \hat{\mathcal{P}}_M$ then $\omega_{(j)} = \omega_{s'}$. The operator $\boldsymbol{D}$ models the effect of the transverse gradient in the Fourier domain.

In parallel to the discussion ending Sec. 3.2, we also restrict the action of $\boldsymbol{D}\boldsymbol{F}$ to the domain $\hat{\mathcal{P}}_M^+$. Consequently, using the operator $\boldsymbol{\Theta}$ (Sec. 3.2), the final linear forward model linking the real FDM to the 2-D NDFT of the discrete RIM $\mathfrak{n}$ reads

$$\boldsymbol{y} = (\boldsymbol{\Theta}\boldsymbol{D}\boldsymbol{F})\mathfrak{n} + \boldsymbol{\eta} \quad \in \mathbb{R}^M. \tag{11}$$

The additional noise $\boldsymbol{\eta} \in \mathbb{R}^M$ integrates the different distortions induced by the numerical computations (*e.g.*, the NFFT inaccuracy when estimating the NDFT $\boldsymbol{F}$, see Appendix B), the model discretization (*e.g.*, the different Riemann approximations), the discrepancy to the paraxial approximation (2), and the actual noise corrupting the observation of $\boldsymbol{z}$. A detailed noise estimation is provided in Sec. 6.

---

[5]This NDFT formulation is strictly equivalent to the one given in [24] where $\delta\tau = \delta r = 1$.

Notice that in the absence of noise ($\boldsymbol{\eta} = 0$), the model (11) could be easily turned into a classical tomographic model where the DC frequency is not observed. Indeed, forgetting the transparent action of $\boldsymbol{\Theta}$, if the frequency origin has been carefully removed from $\hat{\mathcal{P}}_M$, then $\boldsymbol{D}$ is invertible with $\boldsymbol{D}^{-1} = -\frac{i\,\mathfrak{n}_r}{2\pi(\delta r)^2} \operatorname{diag}(\omega_{(1)}^{-1}, \cdots, \omega_{(M)}^{-1})$, and we can solve the common tomographic problem

$$\widetilde{\boldsymbol{y}} := \boldsymbol{D}^{-1}\boldsymbol{y} \ = \ \boldsymbol{F}\,\mathfrak{n}. \tag{12}$$

However, as we present in Sec. 8.1.1, this transformation is not suited to noisy ODT sensing. Even for a simple additive white Gaussian noise $\boldsymbol{\eta}$ (or AWGN), the multiplication by $\boldsymbol{D}^{-1}$ breaks the homoscedasticity of $\boldsymbol{\eta}$, i.e., the variance of each $(\boldsymbol{D}^{-1}\boldsymbol{\eta})_j$ varies with $j \in [M]$. This interferes with common reconstruction techniques used in classical tomography. Obviously, a noise whitening could be realized for stabilizing such methods but at least for AWGN, this strictly amounts to solve directly the model (11).

4   RELATED WORKS

This section describes some recently proposed methods for tomographic reconstruction in the domains of differential phase-contrast tomography and common absorption tomography.

In the domain of differential phase-contrast tomography, we face the problem of recovering the refractive index distribution from phase-shifts measurements. These measurements are composed by the derivative of the refractive index map, inducing the apparition of the *affine* frequency ω when using the FST, as it happens in the ODT sensing model described in Sec. 2. In this domain, Pfeiffer et al. [28] have used the FBP algorithm to reconstruct the refractive index map from a fully covered set of projections. Cong et al. [10, 11] have used different iterative schemes based on the minimization of the TV norm to reconstruct the refractive index distribution over a region of interest. These methods are accurate and provide similar results, but the iterative scheme based on the TV norm has proved to be better than FBP when the amount of acquisitions decreases.

In common Absorption Tomography (AT) we deal with the reconstruction of the absorption index distribution from intensity measurements. As these measurements are directly related to the absorption index, the AT sensing model does not include the *affine* frequency ω. In this domain, several works have exploited sparsity based methods. Most recent works in AT have focused on promoting a small TV norm [30,32]. Sidky et al. [32] use a Lagrangian formulation for the tomographic reconstruction problem, promoting a small TV norm under a Kullback-Leiber data divergence and a positivity constraint. They aim at reconstructing a breast phantom from 60 projections with Poisson distributed noise. For this, they use the primal-dual optimization algorithm proposed by Chambolle et al. [8]. The method results in high quality reconstruction compared to FBP but with a convergence result that is highly dependent on the Lagrangian parameter choosen.

Ritschl et al. [30] use a constrained optimization formulation to reconstruct the absorption index from low amount of clinical data in the presence of metal implants and Gaussian noise. This problem is solved by means of an alternating method that allows then optimizing separately the raw data consistency function and the sparsity cost function, without the need of prior information on the observations. The fast convergence of the method is based on the estimation of the optimization steps. The gradient descent method is used to minimize the TV norm and the consistency term is minimized via an algebraic reconstruction technique. The method is proven to give better results than FBP.

These works have proved that some tomographic applications can be made compressive in the CS sense [6, 12]. They show that it is possible to perform an accurate tomographic image

reconstruction from a number of samples that is smaller than the desired image resolution, when the image is sparse in some basis, *i.e.*, the image expansion in that basis contains only a small number of nonzero coefficients. However, it is important to notice that most of these works have considered Cartesian grids, an approximation of the actual sensing in a polar grid, and plus they attack different problems than ODT, where the sensing model and type of measurement change from one to the other. One important difference between AT and ODT sensing models, relies on the presence of the *affine* frequency as materialized by a diagonal operator $\boldsymbol{D}$; whose impact is analyzed in detail in Sec. 6.3 and Sec. 8.1.1 for perfect and noisy sensing.

5 REFRACTIVE INDEX MAP RECONSTRUCTION

Two methods can be considered for recovering the discrete RIM $\mathfrak{n} \in \mathbb{R}^N$. The first is the Least Square solution, also known as Filtered Back Projection (FBP), and the second is a regularized approach called TV-ℓ_2 minimization. Both procedures are described in detail hereafter.

Least Square Reconstruction

Given any linear sensing model

$$\boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{x} \in \mathbb{R}^M \tag{13}$$

of some image (vector) $\boldsymbol{x} \in \mathbb{R}^N$ by a sensing operator $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ with $M \leqslant N$, the *Least Square* solution $\boldsymbol{x}_{\mathrm{ls}}$ is obtained by solving

$$\boldsymbol{x}_{\mathrm{ls}} = \underset{\boldsymbol{u} \in \mathbb{R}^N}{\arg\min} \|\boldsymbol{u}\|_2 \ \ \text{s.t.} \ \ \boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{u}. \tag{14}$$

This minimization amounts to regularizing the general inverse problem aiming at reconstructing $\boldsymbol{x}$ from (13) by promoting a minimal ℓ_2-norm of the solution. As a matter of fact, $\boldsymbol{x}_{\mathrm{ls}}$ is obtained in a closed form by applying the Moore-Penrose pseudoinverse of $\boldsymbol{\Phi}$ on the observed data $\boldsymbol{y}$, *i.e.*,

$$\boldsymbol{x}_{\mathrm{ls}} = \boldsymbol{\Phi}^\dagger \boldsymbol{y} = \boldsymbol{\Phi}^* (\boldsymbol{\Phi} \boldsymbol{\Phi}^*)^{-1} \boldsymbol{y}.$$

As explained in the previous section, the ODT sensing operator is $\boldsymbol{\Phi} = \boldsymbol{\Theta} \boldsymbol{D} \boldsymbol{F}$, the ODT sensing model is $\boldsymbol{y} = \boldsymbol{\Theta} \boldsymbol{D} \boldsymbol{F} \mathfrak{n}$ and we denote by $\mathfrak{n}_{\mathrm{ls}}$ the corresponding LS solution. Notice that if our forward model followed a classical AT sensing rather than an ODT one, it would be reduced to a simple partial Fourier sampling of $\mathfrak{n}$ by $\boldsymbol{\Phi} = \boldsymbol{\Theta} \boldsymbol{F}$, without the action of $\boldsymbol{D}$.

Notice that the LS solution is equivalent to a discretization of the well known (analytic) Filtered Back Projection [23]. Indeed, for AT and ODT, the expression of FBP can be simplified into a weighting of the 1-D radial frequency measurements (*e.g.*, as obtained with $\boldsymbol{F}^{\mathrm{rad}}$), *i.e.*, a filtering of the spatial measurements, followed by a *backprojection* which amounts to perform a 2-D Fourier inversion of the weighted data extended to the whole frequency domain by zero padding. In AT, the frequency filtering is the so-called "ramp filter" [23], while it can be shown that in ODT it corresponds to a simple Hilbert transform of the affine variations [3, 28]. Henceforth, due to the correspondence between LS and FBP, every solution of the ODT problem obtained with the LS method above will be denoted $\mathfrak{n}_{\mathrm{FBP}}$.

The least square method presents some strong limitations when the model (13) is severely ill-conditioned or when noise corrupts the observation of $\boldsymbol{y}$. For instance, when the ODT sensing is associated to a weak covering of the frequency domain by the grid $\hat{\mathcal{P}}_M$, *i.e.*, for small value of N_θ and $M = N_\theta N_\tau < N = N_0^2$, spurious artifacts appears in the least square estimated RIM $\mathfrak{n}_{\mathrm{FBP}}$. This is better understood in the simplified model where the sensing domain lives also on a Euclidean grid and by once again forgetting the transparent action of $\boldsymbol{\Theta}$. In this case,

Figure 2: (left) RIM TV Model: the gradient of the RIM (represented by arrows) is non-zero only on the interface between the two RIM homogeneous areas $\mathfrak{n}_0$ and $\mathfrak{n}_1$. This induces a small TV norm. (right) The Shepp-Logan phantom, an example of the "Cartoon shape" model.

the NDFT $\boldsymbol{F}$ reduces to a regular DFT $\boldsymbol{F}^{\mathrm{reg}}$ followed by the selection of M frequencies in the N-dimensional Fourier domain, as represented by a restriction matrix $\boldsymbol{R} \in \{0,1\}^{M \times N}$. Therefore, taking into account the correct normalization of the Fourier operators, we have $\boldsymbol{\Phi} = N_0 \, \boldsymbol{DRF}^{\mathrm{reg}}$ and

$$\boldsymbol{\Phi}^*\big(\boldsymbol{\Phi}\boldsymbol{\Phi}^*\big)^{-1}\boldsymbol{y} = \boldsymbol{\Phi}^*\big(\boldsymbol{\Phi}\boldsymbol{\Phi}^*\big)^{-1}\boldsymbol{\Phi}\mathfrak{n} = (\boldsymbol{RF}^{\mathrm{reg}})^*\boldsymbol{RF}^{\mathrm{reg}}\mathfrak{n} = \phi * \mathfrak{n},$$

where ϕ is the Point Spread Function (PSF) of the frequency sampling, i.e., the convolution filter associated to the circulant matrix $(\boldsymbol{RF}^{\mathrm{reg}})^*\boldsymbol{RF}^{\mathrm{reg}}$ (since $\boldsymbol{R}^*\boldsymbol{R}$ is diagonal). While ϕ is a Dirac at $M = N$ involving $\mathfrak{n}_{\mathrm{FBP}} = \mathfrak{n}$, when M is significantly smaller than N, the PSF ϕ is enlarging and contains more and more oscillations. Consequently, the solution $\mathfrak{n}_{\mathrm{FBP}}$ deviates more and more from the pure RIM. This analysis is similar to the characterization of the *dirty map* obtained for instance in radio interferometry by aperture synthesis, which is another tomographic inverse problem [35]. Moreover, as the method does not take the noise into account, the reconstruction is generally smooth and blurred compared to the actual RIM $\mathfrak{n}$ when a significant noise plagues the model (13).

Despite these limitations, since LS/FBP is widely used in tomographic methods, we will use it as a standard to compare with the quality of the regularized reconstruction developed in the following.

TV-ℓ_2 minimization

In order to overcome the limitations of the LS method, we introduce a new reconstruction method which is both less sensitive to unwanted oscillations due to a low density frequency sampling $\hat{\mathcal{P}}_M$ and to additional observational noise $\boldsymbol{\eta}$ in (11).

In particular, since the spatial dimensions in $\mathcal{P}_M$ and in $\mathcal{C}_N$ are expected to be equal, i.e., $N_0 \approx N_\tau$ (Sec. 3.1) we are interested in lowering the density of $\hat{\mathcal{P}}_M$ in the Fourier plane by decreasing the number of angular observations N_θ. In other words, with respect to this reduction, we aim at developing a numerical reconstruction which makes Optical Deflectometric Tomography "*compressive*" in a similar way other compressive imaging devices which, inspired by the Compressed Sensing paradigm [6, 12], reconstruct high resolution images from few (indirect) observations of its content [14, 25, 32, 35]. This ability would lead of course to a significant reduction of the ODT observation time with potential impact, for instance, in fast industrial object quality control relying on this technology.

This objective of compressiveness can only be reached by regularizing the ODT inverse problem by an appropriate prior model on the configuration of the expected RIM $\mathfrak{n}$. Interestingly, the actual RIM of most human made transparent materials (e.g., optical fiber bundles, lenses,

12

optics, ..) is composed by slowly varying areas separated by sharp boundaries (material interfaces) (see Fig. 2-(left)). This can be interpreted with a Bounded Variation (BV) or "Cartoon Shape" model [31] as the typical Shepp-Logan phantom in Fig. 2-(right). Therefore, the inverse problem in (11) can be regularized by promoting a small Total-Variation norm, which in its discrete formulation is defined as [7]

$$\|\mathfrak{n}\|_{TV} := \|\boldsymbol{\nabla}\mathfrak{n}\|_{2,1},$$

where $\|\cdot\|_{2,1}$ is the mixed ℓ_2/ℓ_1 norm defined on any $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2) \in \mathbb{R}^{N \times 2}$ as $\|\boldsymbol{u}\|_{2,1} = \sum_{k=1}^{N}((\boldsymbol{u}_1)_k^2 + (\boldsymbol{u}_2)_k^2)^{1/2}$, and $\boldsymbol{\nabla} : \mathbb{R}^N \to \mathbb{R}^{N \times 2}$ is the (finite difference) gradient operator. Reusing the 2-D coordinates of $\mathfrak{n}$, this operator is defined along each direction as $\boldsymbol{\nabla}\mathfrak{n} = (\boldsymbol{\nabla}_1\mathfrak{n}, \boldsymbol{\nabla}_2\mathfrak{n})$ with $(\boldsymbol{\nabla}_1\mathfrak{n})_{kl} = \mathfrak{n}_{k+1,l} - \mathfrak{n}_{k,l}$ and $(\boldsymbol{\nabla}_2\mathfrak{n})_{kl} = \mathfrak{n}_{k,l+1} - \mathfrak{n}_{k,l}$.

In order to obtain a reconstruction method which is also robust to additive observation noise, we must lighten the strict fidelity constraint implicitly used by the LS method in (14). Therefore, assuming the data corrupted by an additive white Gaussian noise (AWGN), we impose a data fidelity requirement using the ℓ_2-norm, *i.e.*, if $\boldsymbol{\eta} = \boldsymbol{y} - \boldsymbol{\Phi}\mathfrak{n}$ is known to have a bounded norm (or *energy*) $\|\boldsymbol{\eta}\| \leqslant \varepsilon$, we force any reconstruction candidate $\boldsymbol{u}$ to satisfy $\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\| \leqslant \varepsilon$. The way ε can be estimated will be explained in Sec. 6.

Additionally to a fidelity criterion with the observed data, other requirements can be imposed on the reconstruction. First, we can often assume that the reference refractive index $\mathfrak{n}_{\rm r}$ (*i.e.*, the one of some optical fluid surrounding the object) is lower than the object RIM. Second, if the object is completely contained in the field-of-view Ω of the ODT experiment, we can force any candidate RIM $\boldsymbol{u}$ to match $\mathfrak{n}_{\rm r}$ on the boundary of Ω, *i.e.*, imposing $u_k = \mathfrak{n}_{\rm r}$ for all indices k belonging to the border of $\mathcal{C}_N$. Up to a global value shifting the RIM (explained lates in this section), this is equivalent to assume that this map vanishes outside of Ω. These indices are associated to pixels $\boldsymbol{r}_{m,n} \in \mathcal{C}_N$ for which at least one of the two 2-D coordinates is equal to either $-N_0/2$ or $N_0/2 - 1$. The corresponding index set is denoted $\partial\Omega$ for simplicity.

Gathering all these aspects, we could propose the following reconstruction program

$$\mathfrak{n}_{\rm TV-\ell_2} = \arg\min_{\boldsymbol{u} \in \mathbb{R}^N} \|\boldsymbol{u}\|_{TV} \text{ s.t. } \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2 \leqslant \varepsilon, \ \boldsymbol{u} \succeq \mathfrak{n}_{\rm r}, \ \boldsymbol{u}_{\partial\Omega} = \mathfrak{n}_{\rm r}\boldsymbol{1}_{\partial\Omega},$$

denoting by $\boldsymbol{1} \in \mathbb{R}^N$ the vector of ones, *i.e.*, the unit RIM in $\mathcal{C}_N$, and recalling that $\boldsymbol{v}_{\mathcal{A}} = \boldsymbol{R}_{\mathcal{A}}\boldsymbol{v}$ stands for the restriction of the components of $\boldsymbol{v} \in \mathbb{R}^N$ to $\mathcal{A} \subset [N]$.

However, the reconstruction can be slightly simplified by observing that the kernel of the sensing operator $\boldsymbol{\Phi} = \boldsymbol{\Theta}\boldsymbol{D}\boldsymbol{F}$ in ODT contains the set of constant vectors in $\mathbb{R}^N$. This is a consequence of the vanishing affine frequency ω (which mainly defines the action of $\boldsymbol{D}$) on the frequency origin, or more simply, this relies on the occurrence of the RIM gradient in the deflection model (2).

Therefore, a change of variable $\boldsymbol{u} \to \boldsymbol{u} - \mathfrak{n}_{\rm r}\boldsymbol{1}$ does not disturb the previous reconstruction which can be recast as

$$\mathfrak{n}_{\rm TV-\ell_2} = \arg\min_{\boldsymbol{u} \in \mathbb{R}^N} \|\boldsymbol{u}\|_{TV} \text{ s.t. } \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2 \leqslant \varepsilon, \ \boldsymbol{u} \succeq 0, \ \boldsymbol{u}_{\partial\Omega} = 0, \tag{15}$$

remembering that the true RIM estimation is actually $\mathfrak{n}_{\rm TV-\ell_2} + \mathfrak{n}_{\rm r}\boldsymbol{1}$. This last formulation has an interesting property.

Lemma 5.1. *If there is at least one feasible point for the constraints of (15), then the solution of this problem is unique.*

Proof. Using the TV norm definition (and squaring it), the TV-ℓ_2 optimization (15) is equivalent to solve

$$\underset{\boldsymbol{u}\in\mathbb{R}^N}{\arg\min} \|\boldsymbol{\nabla}\boldsymbol{u}\|_{2,1}^2 \ \text{s.t.} \ \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2 \leqslant \varepsilon, \ \boldsymbol{u} \succeq 0, \ \boldsymbol{u}_{\partial\Omega} = 0.$$

We remark also that the kernel of $\boldsymbol{\nabla}$ is the set of constant vectors in $\mathbb{R}^N$ while the kernel of $\boldsymbol{R}_{\partial\Omega}$ (defining the last constraint) is the set of vectors equal to 0 on $\partial\Omega$. Therefore, $\mathrm{Ker}\,\boldsymbol{\nabla}\cap\mathrm{Ker}\,\boldsymbol{R}_{\partial\Omega} = \{\boldsymbol{0}\}$.

Since the domain of $\|\boldsymbol{\nabla}\cdot\|_{2,1}^2$ is $\mathbb{R}^N$, and since we assume at least one feasible point for the constraints of (15), we know that the program has at least one solution. Let us assume that there are two distinct points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ minimizing this optimization and let us denote $\boldsymbol{x}_\lambda = \lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2$ their convex combination for some $0 < \lambda < 1$. By convexity, $\boldsymbol{x}_\lambda$ is also a feasible point of the constraints since both minimizers $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ must satisfy the constraints.

Moreover, $\boldsymbol{R}_{\partial\Omega}\boldsymbol{x}_1 = \boldsymbol{R}_{\partial\Omega}\boldsymbol{x}_2 = 0$ and $\boldsymbol{x}_1 - \boldsymbol{x}_2 \in \mathrm{Ker}\,\boldsymbol{R}_{\partial\Omega}$. Therefore, since $\boldsymbol{x}_1 - \boldsymbol{x}_2 \neq \boldsymbol{0}$, $\boldsymbol{x}_1 - \boldsymbol{x}_2 \notin \mathrm{Ker}\,\boldsymbol{\nabla}$ and $\boldsymbol{\nabla}\boldsymbol{x}_1 \neq \boldsymbol{\nabla}\boldsymbol{x}_2$. By the strict convexity of the function $\varphi(\cdot) = \|\cdot\|_{2,1}^2$, $\varphi(\boldsymbol{\nabla}\boldsymbol{x}_1) = \varphi(\boldsymbol{\nabla}\boldsymbol{x}_2)$ involves

$$\varphi(\boldsymbol{\nabla}\boldsymbol{x}_1) = \lambda\varphi(\boldsymbol{\nabla}\boldsymbol{x}_1) + (1-\lambda)\varphi(\boldsymbol{\nabla}\boldsymbol{x}_2) > \varphi(\boldsymbol{\nabla}\boldsymbol{x}_\lambda),$$

showing that $\boldsymbol{x}_\lambda$ is a better minimizer, which is a contradiction. □

Therefore, we see that the uniqueness is actually reached by the stabilizing condition $\boldsymbol{u}_{\partial\Omega} = \boldsymbol{R}_{\partial\Omega}\boldsymbol{u} = 0$ making the optimization running outside of $\ker\boldsymbol{\nabla}\setminus\{\boldsymbol{0}\}$.

As explained in Section 7, the program (15) can be efficiently solved using proximal methods [9] and operator splitting techniques, like the recently proposed primal-dual algorithm by Chambolle and Pock in [8].

6 NOISE IDENTIFICATION, ESTIMATION AND ANALYSIS

In this section we first discuss about the different sources of noise and how to estimate the noise energy present in the experimental data. Then, we analyze the noise impact in both AT and ODT measurements.

6.1 Noise sources

When a real sensing scenario is being studied, such as the ODT, different sources of noise are present and they have to be considered when determining the global noise energy bound ε in (15).

First, we have the observation noise. Under high light source intensity, the ODT measurements $\boldsymbol{z}$ produced by a Schlieren deflectometer (Fig. 1-(bottom)) are mainly affected by electronic noise such as the CCD thermal noise. This induces a homoscedastic Gaussian noise in the measured deflection angles, *i.e.*, with an homogeneous variance through all the measurements. By computing the 1-D Fourier transform of the ODT measurements using $\boldsymbol{\Theta}\boldsymbol{F}_{\mathrm{rad}}$ in (7) the corresponding noise $\boldsymbol{\eta}_{\mathrm{obs}}$ remains Gaussian [27] in the FDM $\boldsymbol{y}$, *i.e.*, $\boldsymbol{\eta}_{\mathrm{obs}} \sim \mathcal{N}(0, \sigma_{\mathrm{obs}}^2)$.

This noise, defined as the difference between the noisy FDM and the noiseless FDM $\boldsymbol{y}_{\mathrm{true}}$, has an energy that can be bounded using the Chernoff-Hoeffding bound [18]:

$$\|\boldsymbol{y} - \boldsymbol{y}_{\mathrm{true}}\|^2 = \|\boldsymbol{\eta}_{\mathrm{obs}}\|^2 < \varepsilon_{\mathrm{obs}}^2 := \sigma_{\eta_{\mathrm{obs}}}^2(M + c\sqrt{M}),$$

with high probability for $c = \mathcal{O}(1)$.

Second, we have the modeling error that comes with every mathematical discrete representation of a physical continuous system. In the ODT system, this error is due to (i) the paraxial approximation used to formulate (2), (ii) the sampling of the continuous RIM and (iii) the discrete model itself. The modeling noise is related to the difference between the noiseless FDM and the sensing model $\boldsymbol{\Phi}_{\text{true}}\mathfrak{n} = \boldsymbol{\Theta}\boldsymbol{D}\boldsymbol{F}_{\text{true}}\mathfrak{n}$, where $\boldsymbol{F}_{\text{true}}$ performs the exact Polar Fourier Transform. The modeling noise can be computed as:

$$\|\boldsymbol{y}_{\text{true}} - \boldsymbol{\Phi}_{\text{true}}\mathfrak{n}\|^2 = \|\boldsymbol{\eta}_{\text{model}}\|^2 < \varepsilon_{\text{model}}^2.$$

Third, we have the NFFT interpolation noise, given by the mathematical error committed by estimating the polar Fourier Transform with the NFFT algorithm, *i.e.*, the noise $\boldsymbol{\Phi}_{\text{true}}\mathfrak{n} - \boldsymbol{\Phi}\mathfrak{n}$. To determine a bound $\varepsilon_{\text{nfft}}$ on the energy of this error we first estimate the NFFT distortion (*i.e.*, without the action of $\boldsymbol{D}$), defined as the difference between the NFFT polar Fourier Transform $\boldsymbol{F}_{\text{app}}$ and the true NDFT $\boldsymbol{F}$. Theoretically, for any vector $\boldsymbol{f} \in \mathbb{R}^N$, the ℓ_∞-norm of this distortion is bounded as $\|\boldsymbol{F}_{\text{app}}\boldsymbol{f} - \boldsymbol{F}\boldsymbol{f}\|_\infty \leqslant \epsilon\, C(\boldsymbol{f}) = \mathcal{O}(\epsilon\|\boldsymbol{f}\|_1)$, where ϵ controls both the accuracy and the complexity $\mathcal{O}(N \log N/\epsilon)$ of the NFFT [24] (see Appendix B). Assuming that each component of $\boldsymbol{F}_{\text{app}}\boldsymbol{f} - \boldsymbol{F}\boldsymbol{f}$ is iid with a uniform distribution $\mathcal{U}([-C(\boldsymbol{f}), C(\boldsymbol{f})])$, and using the Chernoff-Hoeffding bound [18, 20], we can estimate $\|\boldsymbol{F}_{\text{app}}\boldsymbol{f} - \boldsymbol{F}\boldsymbol{f}\|^2 < \frac{C(\boldsymbol{f})^2}{3}(M + c\sqrt{M})$, with high probability for $c = \mathcal{O}(1)$. Finally, $\varepsilon_{\text{nfft}}$ can be crudely computed as

$$\begin{aligned}
\|\boldsymbol{\Phi}_{\text{true}}\mathfrak{n} - \boldsymbol{\Phi}\mathfrak{n}\| &= \|\boldsymbol{\Theta}\boldsymbol{D}(\boldsymbol{F}_{\text{app}}\mathfrak{n} - \boldsymbol{F}\mathfrak{n})\| \leqslant \|\boldsymbol{\Theta}\boldsymbol{D}\|\|\boldsymbol{F}_{\text{app}}\mathfrak{n} - \boldsymbol{F}\mathfrak{n}\| \\
&= \frac{2\pi(\delta r)^2 \omega_{\max}}{\mathfrak{n}_{\text{r}}} \frac{\epsilon C(\mathfrak{n})}{\sqrt{3}}(M + c\sqrt{M})^{1/2} \\
&\approx \frac{\pi\delta r}{\sqrt{3}\,\mathfrak{n}_{\text{r}}} \epsilon\, C(\mathfrak{n})(M + c\sqrt{M})^{1/2} =: \varepsilon_{\text{nfft}},
\end{aligned}$$

with $\omega_{\max} \approx \frac{1}{2}N_\tau\delta\omega = \frac{1}{2\delta\tau} \approx \frac{1}{2\delta r}$ representing the maximum frequency amplitude in $\hat{\mathcal{P}}_M$. In practice, because of the RIM shift $\mathfrak{n} \to \mathfrak{n} - \mathbf{1}\mathfrak{n}_{\text{r}}$ explained in Sec. 5, we can bound $\|\mathfrak{n}\|_1$ and hence $C(\mathfrak{n})$ with the expected RIM dynamics $\delta\mathfrak{n}$, *i.e.*, $\|\mathfrak{n}\|_1 \leqslant N\delta\mathfrak{n}$, and we adjust ϵ in order to have $\varepsilon_{\text{nfft}}$ much lower than the other sources of noises (mainly ε_{obs}).

Finally, we may also have an error introduced by the instrument calibration, when determining the exact τ and θ associated to the projections. We are going to neglect this error by assuming a pre-calibration process that provides an exact knowledge of these values (see Sec. 8.2).

In conclusion, gathering all the previous noise identifications, we can bound the difference between the actual ODT measurement and the sensing model as follows:

$$\begin{aligned}
\|\boldsymbol{y} - \boldsymbol{\Phi}\mathfrak{n}\| &= \|\boldsymbol{y} - \boldsymbol{y}_{\text{true}} + \boldsymbol{y}_{\text{true}} - \boldsymbol{\Phi}_{\text{true}}\mathfrak{n} + \boldsymbol{\Phi}_{\text{true}}\mathfrak{n} - \boldsymbol{\Phi}\mathfrak{n}\| \\
&\leqslant \|\boldsymbol{y} - \boldsymbol{y}_{\text{true}}\| + \|\boldsymbol{y}_{\text{true}} - \boldsymbol{\Phi}_{\text{true}}\mathfrak{n}\| + \|\boldsymbol{\Phi}_{\text{true}}\mathfrak{n} - \boldsymbol{\Phi}\mathfrak{n}\| \\
&< \varepsilon_{\text{obs}} + \varepsilon_{\text{model}} + \varepsilon_{\text{nfft}} = \varepsilon.
\end{aligned}$$

In the following, we will neglect the error coming from the modeling $\varepsilon_{\text{model}}$, such that only the measurement noise ε_{obs} and the NFFT interpolation noise $\varepsilon_{\text{nfft}}$ are taken into account, with the last one considerably smaller.

6.2 Observation noise estimation for experimental data

When we work with experimental data, we do not have information about the variance of the observation noise σ_{obs}^2 present in the FDM, thus we are not able to estimate the value of ε_{obs}.

For this, we first estimate the variance σ_η of the noise present in the ODT measurements ($\boldsymbol{z}$), denoted by $\eta(\theta, \tau)$. The Robust Median Estimator [13, 34] was used for this purpose. Denoting
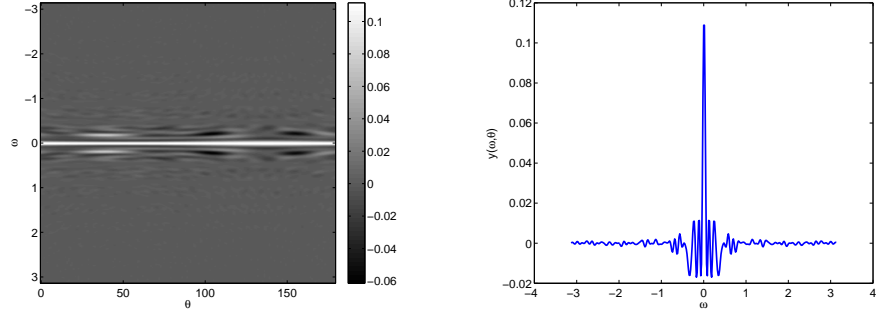
Figure 3: AT Measurement (in Fourier) (left) On the whole grid $\hat{\mathcal{P}}_M$ for $N_\theta = 180$. (right) The slice $\theta = 80°$.
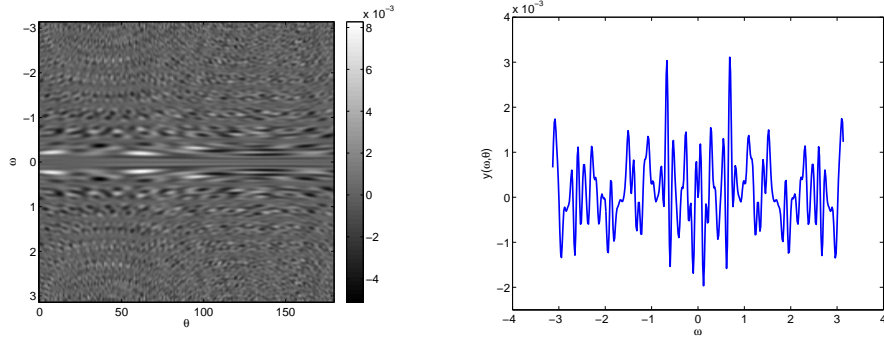

Figure 4: ODT Measurement (in Fourier) (left) On the whole grid $\hat{\mathcal{P}}_M$ for $N_\theta = 180$. (right) The slice $\theta = 80°$.

by $\Psi_{\eta,\theta}(\tau) = \Delta_\eta(\tau, \theta)$ the measured signal at one specific θ, the median estimator is applied as follows:

$$\sigma_e^2(\theta) = \frac{\text{median}_\tau \, |h * \Psi_{\eta,\theta}(\tau)|}{0.6745 \, \|h\|},$$

where $h(\tau)$ is a high-pass filter, $*$ represents the discrete convolution on τ and the median is taken for all the values of $|h * \Psi_{\eta,\theta}(\tau)|$. The resulting $\sigma_e^2(\theta)$ is the estimated variance of the observation noise present in the projections $\boldsymbol{z}$. As the Fourier transform preserves the variance unchanged, the variance of the noise present in the FDM is the same as the one of the observations, i.e., $\sigma_{\text{obs}}^2(\theta) = \sigma_e^2(\theta)$.

6.3   AT vs ODT Noise

As we have seen in Sec. 3.3, the main difference between the AT and ODT problems is the appearance of the diagonal operator $\boldsymbol{D}$ in the last one. We will now analyze the impact of additive white Gaussian noise on the measurements regarding the presence of this operator.

In Fig. 3 and Fig. 4 we show the magnitude of the Fourier measurements in AT and ODT, respectively, corresponding to the acquisition of a section of the fibers bundle (see Fig. 5-(left)). For the class of images we are interested in, we can notice than in AT the magnitude presents a peak around $\omega = 0$ and then decreases significantly when the distance to the center increases, tending to zero in the borders (see Fig. 3). Whereas in ODT, the presence of operator $\boldsymbol{D}$ makes the image intensity to be quite spread through all the pixels (see Fig. 4). This has a direct impact on the reconstruction when the measurement is affected by additive Gaussian noise. As the noise spreads evenly through the image, the pixels that are not around $\omega = 0$ will be more affected in the AT model because their intensity is significantly lower.

7   NUMERICAL METHODS

This section provides first an overview of the primal-dual algorithm we use for solving (15) and hence recovering the RIM from the noisy FDM. Then, the algorithm is generalized into a product space optimization that allows the minimization of more than two convex functions. Finally, we conclude the section by showing how to use this generalization to solve our ODT problem.

7.1   The Primal-Dual Chambolle-Pock Algorithm

The Chambolle-Pock (CP) algorithm (Algorithm 1 in [8]) is an efficient, robust and flexible algorithm that allows to solve minimization problems of the form:

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \ F(\boldsymbol{K}\boldsymbol{x}) + G(\boldsymbol{x}), \tag{16}$$

for a linear operator $\boldsymbol{K} : \mathbb{R}^N \to \mathbb{R}^W$ and any variable $\boldsymbol{x} \in \mathbb{R}^N$. The functions F and G belong to the functional sets $\Gamma^0(\mathbb{R}^W)$ and $\Gamma^0(\mathbb{R}^N)$, respectively.

In short, CP solves the primal problem described above simultaneously with its dual problem, until the difference between their objective functions – the primal-dual gap – is zero.

For any variable $\boldsymbol{v} \in \mathbb{R}^W$, the primal-dual optimization can be formulated as the following saddle-point problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \max_{\boldsymbol{v}\in\mathbb{R}^W} \langle \boldsymbol{K}\boldsymbol{x}, \boldsymbol{v} \rangle + G(\boldsymbol{x}) - F^*(\boldsymbol{v}),$$

where F^* is the convex conjugate of function F provided by the Legendre transform $F^*(\boldsymbol{v}) = \max_{\bar{\boldsymbol{v}}\in\mathbb{R}^W} \langle \boldsymbol{v}, \bar{\boldsymbol{v}} \rangle - F(\bar{\boldsymbol{v}})$.

Using the Legendre transform we obtain the primal version described in (16) and also the dual version as follows:

$$\max_{\boldsymbol{v}\in\mathbb{R}^W} -F^*(\boldsymbol{v}) - G^*(-\boldsymbol{K}^*\boldsymbol{v}), \tag{17}$$

where $\boldsymbol{K}^*$ is the exact adjoint of the operator $\boldsymbol{K}$, such that $\langle \boldsymbol{K}\boldsymbol{x}, \boldsymbol{v} \rangle = \langle \boldsymbol{x}, \boldsymbol{K}^*\boldsymbol{v} \rangle$.

The CP algorithm is defined by the following iterations:

$$\begin{cases} \boldsymbol{v}^{(k+1)} & = \mathrm{prox}_{\nu F^*}(\boldsymbol{v}^{(k)} + \nu \boldsymbol{K}\bar{\boldsymbol{x}}^{(k)}), \\ \boldsymbol{x}^{(k+1)} & = \mathrm{prox}_{\mu G}(\boldsymbol{x}^{(k)} - \mu \boldsymbol{K}^*\boldsymbol{v}^{(k+1)}), \\ \bar{\boldsymbol{x}}^{(k+1)} & = \boldsymbol{x}^{(k+1)} + \vartheta(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}). \end{cases} \tag{18}$$

The quantity prox_f denotes the *proximal operator* of a convex function $f \in \Gamma^0(\mathcal{V})$ for a certain finite dimensional vector space $\mathcal{V}$ [9, 26]. This operator is defined as:

$$\mathrm{prox}_f \boldsymbol{\zeta} \ := \ \underset{\boldsymbol{\zeta}'\in\mathcal{V}}{\arg\min} f(\boldsymbol{\zeta}') + \tfrac{1}{2}\|\boldsymbol{\zeta} - \boldsymbol{\zeta}'\|^2, \qquad \boldsymbol{\zeta} \in \mathcal{V}.$$

The proximal operator admits the use of non-smooth convex functions as the TV norm, making the algorithm suitable to solve the TV-ℓ_2 problem described in Sec. 5.

Most numerical methods require operator $\boldsymbol{K}$ being in a tight frame, which is not the case for our sensing operator $\boldsymbol{\Phi}$. The CP algorithm reduces the convergence requirements, since we only need to tune the step sizes μ and ν such that the condition $\mu\nu\|\boldsymbol{K}\|^2 < 1$ is true for any operator $\boldsymbol{K}$.

7.2 Primal-Dual Gap

The primal-dual gap, defined as the difference between the primal and dual objective functions [5], helps us to perform a pratical analysis of the algorithm convergence.

Setting P as the cost of the primal problem from (16) and D as the cost of the dual problem from (17), we can define the estimated primal-dual gap as $\text{PDG} = P - D$. While iterating, the primal objective function is greater than the dual objective function and when approching to the optimal solution, the difference between these objective functions tends to zero.

If $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{v}})$ is the output of the CP algorithm, we are interested in having the primal-dual gap

$$\text{PDG}(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{v}}) = F(\boldsymbol{K}\widetilde{\boldsymbol{x}}) + G(\widetilde{\boldsymbol{x}}) + F^*(\widetilde{\boldsymbol{v}}) + G^*(-\boldsymbol{K}^*\widetilde{\boldsymbol{v}}), \tag{19}$$

close to zero. When the PDG is exactly zero, the strong duality holds and the solution $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{v}})$ is optimal [5].

7.3 Product Space Optimization

In this paper, we are interested in minimization problems containing more than two convex functions. In particular, we aim at solving the general optimization

$$\min_{\boldsymbol{x} \in \mathbb{R}^N} \sum_{j=1}^{p} F_j(\boldsymbol{K}_j \boldsymbol{x}) + G(\boldsymbol{x}),$$

with $\boldsymbol{K}_j : \mathbb{R}^N \to \mathbb{R}^{W_j}$ and $p + 1$ the number of convex functions. Such a problem does not allow the direct use of the CP algorithm as described before. However, it is easy to adapt it by considering a p-times expanded optimization space $\mathbb{R}^{pN}$. This space is composed of $\boldsymbol{t} = (\boldsymbol{t}_1^T, \cdots, \boldsymbol{t}_p^T)^T \in \mathbb{R}^{pN}$, with $\boldsymbol{t}_j \in \mathbb{R}^N$. In this context, we define $p - 1$ bisector planes $\Pi_{1,j} = \{\boldsymbol{t} \in \mathbb{R}^{pN} : \boldsymbol{t}_1 = \boldsymbol{t}_j, \ 2 \leqslant j \leqslant p\}$ in order to work with the following equivalent primal problem:

$$\min_{\boldsymbol{t} \in \mathbb{R}^{pN}} \sum_{j=1}^{p} F_j(\boldsymbol{K}_j \boldsymbol{t}_j) + \sum_{j=2}^{p} \imath_{\Pi_{1,j}}(\boldsymbol{t}) + H(\boldsymbol{t}_1). \tag{20}$$

The CP-shape (16) is thus recovered by working in this bigger space $\mathbb{R}^{pN}$ and by setting $F(\boldsymbol{s}) = \sum_{j=1}^{p} F_j(\boldsymbol{s}_j)$, with $\boldsymbol{s} = (\boldsymbol{s}_1^T, \cdots, \boldsymbol{s}_p^T)^T \in \mathbb{R}^{W = \sum_j W_j}$ and $\boldsymbol{s}_j \in \mathbb{R}^{W_j}$, $\boldsymbol{K} = \text{diag}(\boldsymbol{K}_1, \cdots, \boldsymbol{K}_p) \in \mathbb{R}^{W \times pN}$ and $G(\boldsymbol{t}) = \sum_{j=2}^{p} \imath_{\Pi_{1,j}}(\boldsymbol{t}) + H(\boldsymbol{t}_1)$.

In this expanded optimization space, the equivalent dual problem is written (see Appendix C.1):

$$\max_{\boldsymbol{s} \in \mathbb{R}^W} \ -\sum_{j=1}^{p} F_j^*(\boldsymbol{s}_j) - H^*\left(-\sum_{j=1}^{p} \boldsymbol{K}_j^* \boldsymbol{s}_j\right).$$

For the functions described above, it is easy to see that for any $\nu > 0$ and $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1^T, \cdots, \boldsymbol{\zeta}_p^T)^T \in \mathbb{R}^W$ we have:

$$\text{prox}_{\nu F^*} \boldsymbol{\zeta} = \begin{pmatrix} \text{prox}_{\nu F_1^*} \boldsymbol{\zeta}_1 \\ \vdots \\ \text{prox}_{\nu F_p^*} \boldsymbol{\zeta}_p \end{pmatrix},$$

and, for any $\mu > 0$ we have (more details in Appendix D.1):

$$\text{prox}_{\mu G} \boldsymbol{\zeta} = (\mathbb{I}^N, \cdots, \mathbb{I}^N)^T \, \text{prox}_{\frac{\mu}{p} H}\left(\tfrac{1}{p}\textstyle\sum_j \boldsymbol{\zeta}_j\right).$$

7.4 ODT numerical reconstruction

Now we need to transform our TV-ℓ_2 problem into an expanded form in order to use the CP algorithm. Having two constraints, the optimization space needs to be expanded by $p = 2$. Using (20), we can reformulate the primal problem from (15) as

$$\min_{\boldsymbol{t}=(\boldsymbol{t}_1,\boldsymbol{t}_2)\in\mathbb{R}^{2N}} \|\boldsymbol{\nabla}\boldsymbol{t}_1\|_{2,1} + \imath_{\mathcal{C}}(\boldsymbol{\Phi}\boldsymbol{t}_2) + \imath_{\mathcal{P}_0}(\boldsymbol{t}_1) + \imath_{\Pi_{1,2}}(\boldsymbol{t}), \tag{21}$$

where $\mathcal{C} = \{\boldsymbol{v}\in\mathbb{R}^M : \|\boldsymbol{y}-\boldsymbol{v}\| \leqslant \varepsilon\}$ and $\mathcal{P}_0 = \{\boldsymbol{u}\in\mathbb{R}^N : u_i \geqslant 0 \text{ if } i\in\text{int } \Omega; \; u_i = 0 \text{ if } i\in\partial\Omega\}$. We show easily that (21) has the shape of (20) with $F_1(\boldsymbol{s}_1) = \|\boldsymbol{s}_1\|_{2,1}$ for $\boldsymbol{s}_1\in\mathbb{R}^{N\times 2}\simeq\mathbb{R}^{2N}$, $F_2(\boldsymbol{s}_2) = \imath_{\mathcal{C}}(\boldsymbol{s}_2)$ for $\boldsymbol{s}_2\in\mathbb{R}^M$, $H(\boldsymbol{t}_1) = \imath_{\mathcal{P}_0}(\boldsymbol{t}_1)$, $\boldsymbol{K}_1 = \boldsymbol{\nabla}$ and $\boldsymbol{K}_2 = \boldsymbol{\Phi} = \boldsymbol{\Theta}\boldsymbol{D}\boldsymbol{F}\in\mathbb{R}^{M\times N}$.

For building the dual problem, we need the conjugate functions of F_1, F_2 and H, which are easily computed using the Legendre transform. As a matter of fact, F_1^* is the indicator function onto the convex set $\mathcal{Q} = \{\boldsymbol{q} = (\boldsymbol{q}_1,\boldsymbol{q}_2)\in\mathbb{R}^{N\times 2} : \|\boldsymbol{q}\|_{2,\infty} \leqslant 1\}$ with the mixed ℓ_∞/ℓ_2-norm defined as $\|\boldsymbol{q}\|_{2,\infty} = \max_k \sqrt{(\boldsymbol{q}_1)_k^2 + (\boldsymbol{q}_2)_k^2}$ [8]. The conjugate function F_2^* is computed as (see Appendix. C.2)

$$F_2^*(\boldsymbol{s}_2) = \imath_{\mathcal{C}}^*(\boldsymbol{s}_2) = (\boldsymbol{s}_2)^T\boldsymbol{y} + \varepsilon\|\boldsymbol{s}_2\|,$$

while the convex conjugate of H is simply $H^*(\boldsymbol{s}) = \imath_{\mathcal{P}_0}(-\boldsymbol{s})$.

The dual optimization problem is thus defined as:

$$\max_{\boldsymbol{s}\in\mathbb{R}^{2N+M}} -\imath_{\mathcal{Q}}(\boldsymbol{s}_1) - \langle\boldsymbol{s}_2,\boldsymbol{y}\rangle - \varepsilon\|\boldsymbol{s}_2\| - \imath_{\mathcal{P}_0}(\boldsymbol{\nabla}^*\boldsymbol{s}_1 + \boldsymbol{\Phi}^*\boldsymbol{s}_2).$$

In order to apply (18), we must compute the proximal operators of F_1^*, F_2^* and H. The one of F_1^* is given by [8]

$$(\text{prox}_{\nu F_1^*}\boldsymbol{\zeta})_k = \frac{\big((\boldsymbol{\zeta}_1)_k,(\boldsymbol{\zeta}_2)_k\big)}{\max(1,\sqrt{(\boldsymbol{\zeta}_1)_k^2+(\boldsymbol{\zeta}_2)_k^2})}, \quad \boldsymbol{\zeta} = (\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2)\in\mathbb{R}^{N\times 2}.$$

The proximal operator of F_2^* is determined via the proximal operator of F_2 by means of the conjugation property defined in [9]:

$$\text{prox}_{\nu F_2^*}\boldsymbol{s}_2 = \boldsymbol{s}_2 - \nu\,\text{prox}_{\frac{1}{\nu}F_2}\frac{1}{\nu}\boldsymbol{s}_2.$$

The proximal operator of F_2 is given by the projection onto the convex set $\mathcal{C}$:

$$\text{prox}_{\frac{1}{\nu}F_2}\boldsymbol{s}_2 = \boldsymbol{y} + (\boldsymbol{s}_2 - \boldsymbol{y})\min\left(1, \frac{\varepsilon}{\|\boldsymbol{s}_2-\boldsymbol{y}\|}\right).$$

The proximal operator of the function H represents a projection onto the positive orthant with zero borders:

$$\text{prox}_{\frac{\mu}{2}H}\boldsymbol{\xi} = \text{proj}_{\mathcal{P}_0}\boldsymbol{\xi} = \begin{cases} (\xi_i)_+ & \text{if } i\in\text{int }\Omega, \\ 0 & \text{if } i\in\partial\Omega, \end{cases}, \quad \boldsymbol{\xi}\in\mathbb{R}^N.$$

Finally, making use of the above computations and taking $\vartheta = 1$, the CP algorithm applied to our TV-ℓ_2 problem in ODT can be reduced to (see Appendix D.2)

$$\begin{cases} \boldsymbol{s}_1^{(k+1)} &= \text{prox}_{\nu F_1^*}\big(\boldsymbol{s}_1^{(k)} + \nu\boldsymbol{\nabla}\bar{\boldsymbol{x}}^{(k)}\big), \\ \boldsymbol{s}_2^{(k+1)} &= \text{prox}_{\nu F_2^*}\big(\boldsymbol{s}_2^{(k)} + \nu\boldsymbol{\Phi}\bar{\boldsymbol{x}}^{(k)}\big), \\ \boldsymbol{x}^{(k+1)} &= \text{proj}_{\mathcal{P}_0}\big(\boldsymbol{x}^{(k)} - \frac{\mu}{2}(\boldsymbol{\nabla}^*\boldsymbol{s}_1^{(k+1)} + \boldsymbol{\Phi}^*\boldsymbol{s}_2^{(k+1)})\big), \\ \bar{\boldsymbol{x}}^{(k+1)} &= 2\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}. \end{cases} \tag{22}$$

In our experiments, the $\boldsymbol{x}^{(0)}$, $\bar{\boldsymbol{x}}^{(0)}$, $\boldsymbol{s}_1^{(0)}$ and $\boldsymbol{s}_2^{(0)}$ are initialized to zero vectors.

In order to guarantee the convergence of the algorithm, $i.e.$, to ensure that $\boldsymbol{x}^{(k)}$ converges to the solution of (15) when k increases, we need to set μ and ν such that $\mu\nu\|\boldsymbol{K}\|^2 < 1$. We are interested in giving more weight to the regularization term (TV-norm) in (15). As the TV-norm is contained in function F, we need to assign to the parameter ν in CP a higher value compared to μ. Taking $\nu = 10^3\mu$ provides thus more rapid convergence. The induced norm of the operator ($\|\boldsymbol{K}\|$) was computed as explained in [32] using the standard power iteration algorithm to calculate the largest singular value of the associated matrix $\boldsymbol{K}$.

The algorithm presented in (22) stops when it achieves a stable behavior, $i.e.$, when $\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|/\|\boldsymbol{x}^{(k)}\| \leqslant$ Th. The threshold Th is defined for ODT in the next section. In parallel, we analyze the convergence of the algorithm by means of the primal-dual gap as described in (19). If $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{s}})$ is the output vector of the CP algorithm (22), the primal-dual gap reads

$$\mathrm{PDG}(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{s}}) = \|\boldsymbol{\nabla}\widetilde{\boldsymbol{x}}\|_{2,1} + \imath_{\mathcal{C}}(\boldsymbol{\Phi}\widetilde{\boldsymbol{x}}) + \imath_{\mathcal{P}}(\widetilde{\boldsymbol{x}}) + \imath_{\mathcal{Q}}(\widetilde{\boldsymbol{s}}_1)$$
$$+ \ \langle\widetilde{\boldsymbol{s}}_2, \boldsymbol{y}\rangle + \varepsilon\|\widetilde{\boldsymbol{s}}_2\| + \imath_{\mathcal{P}_0}(\boldsymbol{\nabla}^*\widetilde{\boldsymbol{s}}_1 + \boldsymbol{\Phi}^*\widetilde{\boldsymbol{s}}_2).$$

Due to the presence of the indicator functions, the primal-dual gap will converge to zero only when their value is zero, $i.e.$, when the argument of the indicator function belongs to the corresponding convex set. We will therefore evaluate a partial primal-dual gap

$$\mathrm{pPDG}(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{s}}) = \|\boldsymbol{\nabla}\widetilde{\boldsymbol{x}}\|_{2,1} + \langle\widetilde{\boldsymbol{s}}_2, \boldsymbol{y}\rangle + \varepsilon\|\widetilde{\boldsymbol{s}}_2\|, \tag{23}$$

and, in parallel, we monitor the value of the different indicator functions, such that we comply with the following conditions:

$$\boldsymbol{\Phi}\widetilde{\boldsymbol{x}} \in \mathcal{C}, \tag{24a}$$
$$\widetilde{\boldsymbol{x}} \in \mathcal{P}_0, \tag{24b}$$
$$\widetilde{\boldsymbol{s}}_1 \in \mathcal{Q}, \tag{24c}$$
$$(\boldsymbol{\nabla}^*\widetilde{\boldsymbol{s}}_1 + \boldsymbol{\Phi}^*\widetilde{\boldsymbol{s}}_2) \in \mathcal{P}_0. \tag{24d}$$

8 EXPERIMENTS

In this section, the ODT reconstruction is first compared to the common tomographic (AT) reconstruction using the FBP method. Then, the proposed regularized reconstruction (TV-ℓ_2) is compared with the FBP method on synthetic and experimental ODT data.

8.1 Synthetic Data

Three kinds of discrete synthetic 2-D RIM are selected to test the reconstruction. They are defined on a 256×256 pixel grid ($N = 256^2$). In the first object, the RIM ($\mathfrak{n}$) simulates a 2-D section of a bundle of 10 fibers of radius 8 pixels each, immersed in an optical fluid (the background). The two media have a refractive index difference of $\delta\mathfrak{n} = 12 \times 10^{-3}$ (see Fig. 5-(left)). The second object consists in a homogeneous ball centered in the pixel $(154, 154)$ with a radius of 60 pixels, immersed in a liquid with $\delta\mathfrak{n} = 2.8 \times 10^{-3}$ (see Fig. 5-(right)). These two objects were selected in correspondence to the available experimental data we use for reconstruction later in this section. The third object is the well-known Shepp-Logan phantom (see Fig. 2-(right)), which is a more complex image in a "Cartoon-shape" model.

The measurements were simulated according to (11) by means of the operator $\boldsymbol{\Phi}$, and then, additive white Gaussian noise $\boldsymbol{\eta}_{\mathrm{obs}} \sim_{\mathrm{iid}} \mathcal{N}(0, \sigma_{\eta_{\mathrm{obs}}}^2)$ is added in order to simulate a realistic ODT scenario.
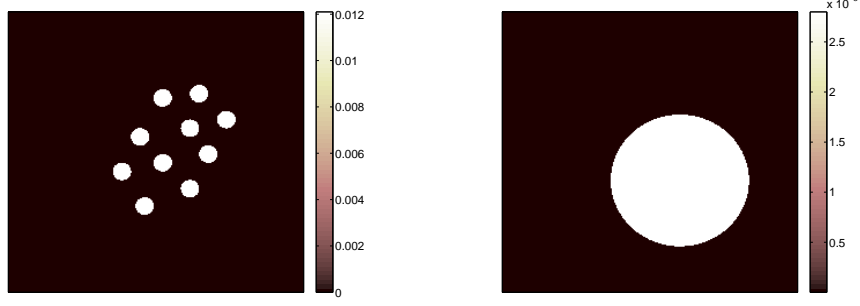
Figure 5: Realistic refractive index map: (left) Synthetic fibers bundle and (right) Synthetic ball.

The operator $\boldsymbol{\Phi}$ is defined as $\boldsymbol{\Phi}_\epsilon = \boldsymbol{\Theta}\boldsymbol{D}\boldsymbol{F}_\epsilon$, with ϵ representing the distorsion of F_ϵ regarding a true operator $\boldsymbol{F}_{\text{true}}$ that would provide the actual NDFT. As discussed in Sec. 3.3, the NDFT computational time is inversely proportional to this parameter ϵ in $\mathcal{O}(N \log(N/\epsilon))$. Therefore, we need to do a compromise between an accurate and an efficient computation of the NDFT. For this reason we use two different operators: (i) an accurate and high dimensional operator $\boldsymbol{\Phi}_{\epsilon_0} = \boldsymbol{\Theta}\boldsymbol{D}\boldsymbol{F}_{\epsilon_0}$ for the acquisition, with a small $\epsilon_0 = 10^{-14}$; and (ii) a less accurate but lower dimensional operator $\boldsymbol{\Phi}_{\epsilon_1} = \boldsymbol{\Theta}\boldsymbol{D}\boldsymbol{F}_{\epsilon_1}$ for the reconstruction, with $\epsilon_1 > \epsilon_0$. The error caused for using a higher ϵ for the reconstruction is taken into account in $\varepsilon_{\text{nfft}}$ (see Sec. 6).

For each object, the ODT measurements are obtained with $N_\tau = 367$ according to a varying number of orientations N_θ, which allows to analyze the compressiveness of the reconstruction method. In this synthetic experiment, the orientations θ are taken in $[0, \pi)$ so that $N_\theta = 360$ corresponds to two orientations per degree. Hereafter, we consider this last situation, i.e., $\delta\theta = \pi/360$ as a "full observation" scenario since, given the considered RIM resolutions, the discrete frequency plane is almost fully covered in this case. More generally, we say that a given orientation number N_θ is associated to $(100 \frac{N_\theta}{N_{\text{full}}})\%$ of the full coverage, N_{full} being the number of orientations for having $\delta\theta = \pi/360$.

The reconstruction robustness with respect to the noise level has been considered by taking σ_η so that the Measurement SNR (MSNR), as measured by MSNR $= 20 \log_{10} \|\boldsymbol{\Delta}\|/\|\boldsymbol{\eta}\|$, is taken in $\{10\,\text{dB}, 20\,\text{dB}, \infty\}$. This last case with MSNR close to $+\infty$ corresponds to the noiseless scenario, where no Gaussian noise is added, only the NFFT interpolation error ($\boldsymbol{\eta}_{\text{nfft}}$) is taken into account. This actually provides a high MSNR value around $270\,\text{dB}$.

The reconstruction quality of $\widetilde{\mathfrak{n}} \in \{\mathfrak{n}_{\text{FBP}}, \mathfrak{n}_{\text{TV}-\ell_2}\}$ is measured using the Reconstruction SNR (RSNR) measured by RSNR $= 20 \log_{10} \|\mathfrak{n}\|/\|\mathfrak{n} - \widetilde{\mathfrak{n}}\|$.

8.1.1   AT vs ODT

In order to assess numerically the impact of operator $\boldsymbol{D}$ in ODT, we compare the RSNR between AT and ODT in similar noisy acquisition scenario. The comparison is made using the FBP algorithm, commonly applied in tomographic reconstructions. We analyze the impact of the affine frequency ω, present in ODT, via the compressiveness and noise robustness. For this, we focus on the reconstruction of the bundle of fibers for different number of orientations $N_\theta \in \{4, 18, 36, 90, 180, 360\}$. The results are depicted in Fig. 6.

In Fig. 6-(left) we can see that, when no Gaussian noise is added, we obtain similar RSNR for both AT and ODT. The impact of the parameter ω is evident only in the convergence time, causing the ODT reconstruction to be 4 times slower than the AT one. However, when we add Gaussian noise in such a way that both data have a MSNR $= 20\,\text{dB}$, the AT reconstruction presents a fast degradation while the ODT reconstruction remains almost unaffected by the
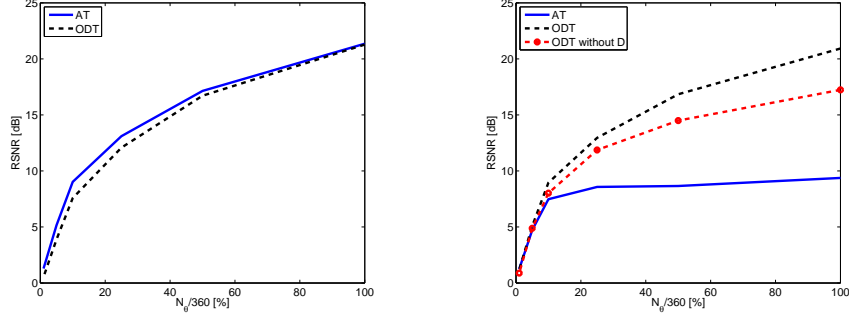
Figure 6: Absorption Tomography vs Deflectometric Tomography for different number of orientations N_θ with (left) MSNR $= \infty$ and (right) MSNR $= 20$ dB.

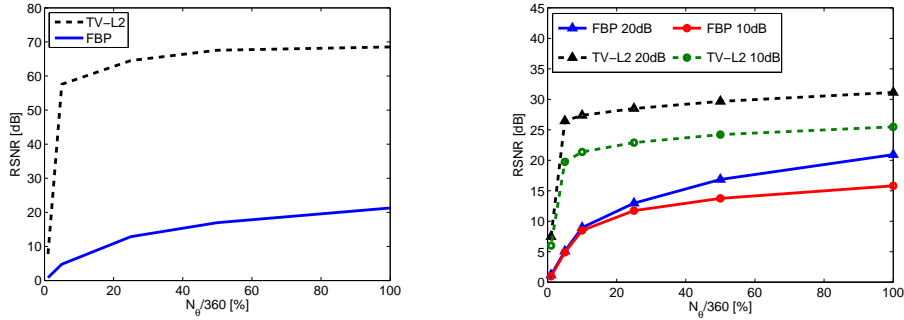Figure 7: FBP vs. TV-ℓ_2 for different number of orientations N_θ with (left) MSNR $= \infty$ and (right) MSNR $= 20$ dB and MSNR $= 10$ dB.

noise (see Fig. 6-(right)). These results corroborate the discussion in Sec. 6.3.

Following the discussion from Sec. 3.3, we analyze now the reconstruction of the RIM using a simplified ODT sensing model that is close to a classical tomographic model (12). In Fig. 6-(right) we show a third curve that corresponds to the RIM reconstruction from a noisy ODT sensing where the diagonal operator $\boldsymbol{D}$ has been removed (ODT without $\boldsymbol{D}$) as in (12). The results were obtained using the FBP algorithm and for a MSNR $= 20$ dB. As it was expected, when removing the operator $\boldsymbol{D}$ from the ODT sensing model, the reconstruction quality decreases significantly compared to the results obtained with the complete ODT sensing model (11). Moreover, the regularized formulation TV-ℓ_2 cannot be used for this ODT reconstruction because the noise is then heteroscedastic.

8.1.2 TV-ℓ_2 Reconstruction method

The TV-ℓ_2 reconstruction is compared with the common FBP method. The reconstruction quality is investigated with respect to compressiveness and noise robustness. Fig. 7 presents comparison graphs of FBP and TV-ℓ_2 showing the RSNR vs the number of orientations $N_\theta \in \{4, 18, 36, 90, 180, 360\}$ for the three noise scenarios. These results correspond to the reconstruction of the bundle of fibers for Th $= 10^{-5}$.

In Fig. 7-(top left) we present the scenario without added noise, $i.e.$, MSNR $= \infty$. We can see that for a full coverage, $i.e.$, $N_\theta = 360$, as the TV-ℓ_2 method takes into account the small noise coming from the NFFT interpolation error, it provides a very good reconstruction that outperforms by 47 dB the FBP reconstruction quality.

The FBP method degrades rapidly when the problem is ill-posed, $i.e.$, when the projections space is not fully covered, whereas the TV-ℓ_2 method maintains a high performance. By pro-

(a) $\mathfrak{n}_{\text{FBP}}$. 5%. RSNR = 4.78 dB. (b) $\mathfrak{n}_{\text{TV}-\ell_2}$. 5%. RSNR = 57.59 dB.

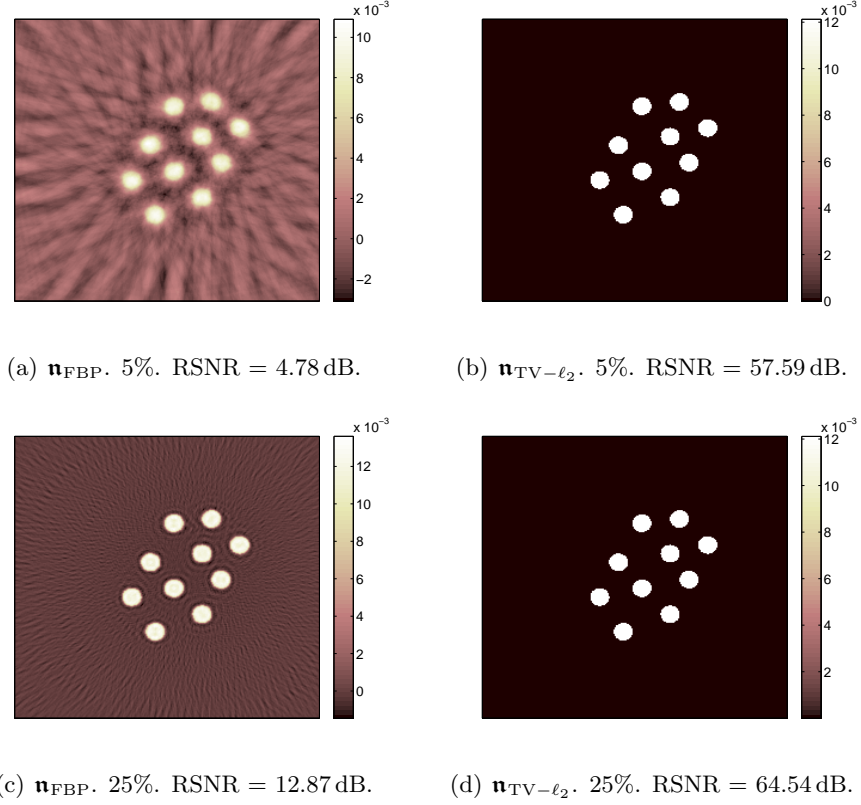(c) $\mathfrak{n}_{\text{FBP}}$. 25%. RSNR = 12.87 dB. (d) $\mathfrak{n}_{\text{TV}-\ell_2}$. 25%. RSNR = 64.54 dB.

Figure 8: Reconstruction images using FBP and TV-ℓ_2 reconstruction methods for MSNR = ∞ and different number of orientations. In the left column we have the FBP reconstruction results for (a) $N_\theta = 18$ and (c) $N_\theta = 90$. In the right column we have the TV-ℓ_2 reconstruction results for (b) $N_\theta = 18$ and (d) $N_\theta = 90$.

moting a small TV-norm, the regularized method presents high compressiveness, as it can be observed in the graph where a high reconstruction quality is still achieved at only 5% of 360 incident angles, obtaining a gain of 53 dB over FBP. Although the performance of the algorithm decreases significantly for a coverage of 1%, it still provides a higher reconstruction quality than FBP.

The high compressiveness properties of the TV-ℓ_2 method are preserved when we add Gaussian noise. We are able to obtain good quality images even for a compressive and highly noisy sensing. With TV-ℓ_2, at a MSNR = 10 dB, we get a RSNR of 16 dB for a 5% radial coverage compared to 4 dB for FBP. However, we can notice how the reconstruction quality of TV-ℓ_2 diminishes with respect to the noiseless scenario, whereas FBP is less affected by the noise.

Up to now, we have shown the behavior of the algorithms for different noise scenarios and compressiveness ratios. Let us now observe a few reconstructed images in order to appreciate their visual quality and the difference with the ground truth for both methods. Fig. 8 presents the resulting images when reconstructing the bundle of fibers in a noiseless scenario and for $N_\theta = \{18, 90\}$, which represents, respectively, a coverage of 5% and 25% of the frequency plane. The algorithm is set to stop when $\texttt{Th} = 10^{-5}$ is reached. In Table 1 we present the number of iterations and the time spent to obtain the results showed in Fig. 8.

We can notice how the TV-ℓ_2 method preserves the image dynamics even for 5% of coverage, while FBP provides images with implausible negative values. We can also observe that some artifacts appear in the FBP results when the problem becomes more and more ill-posed.

About the numerical complexities shown in Table 1 we can notice that, to reach the same

	$N_\theta = 18(5\%)$		$N_\theta = 90(25\%)$	
	# iter	time	# iter	time
TV-ℓ_2	33270	4h31'	10590	1h39'
FBP	18930	2h07'	10220	1h04'

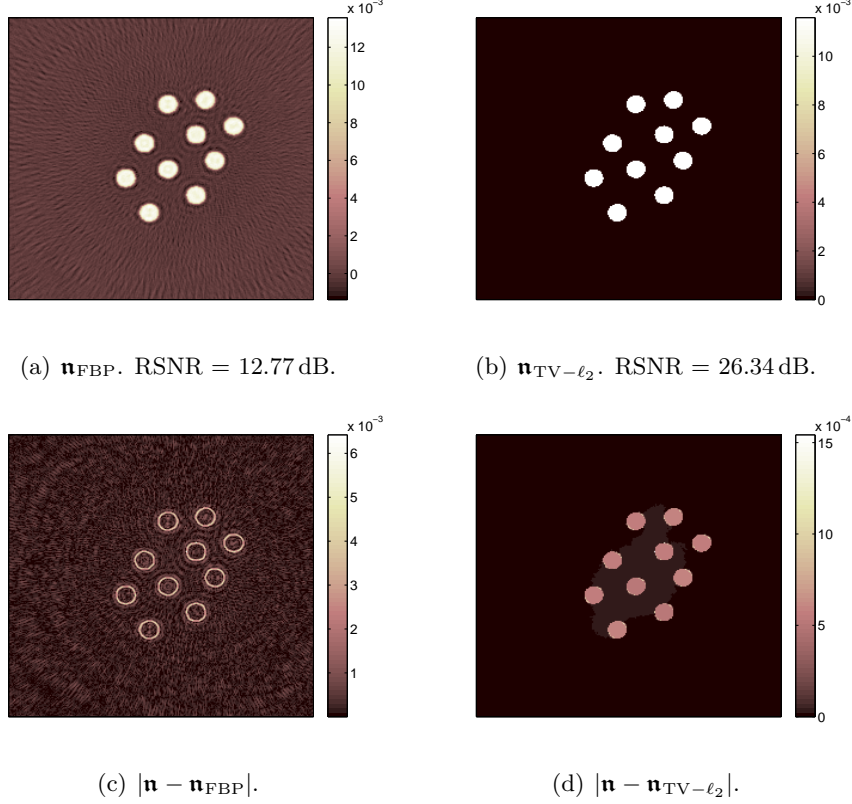Table 1: Number of iterations and time for $N_\theta = \{18, 90\}$ and MSNR $= \infty$.

(a) $\mathfrak{n}_{\mathrm{FBP}}$. RSNR $= 12.77\,\mathrm{dB}$.

(b) $\mathfrak{n}_{\mathrm{TV}-\ell_2}$. RSNR $= 26.34\,\mathrm{dB}$.

(c) $|\mathfrak{n} - \mathfrak{n}_{\mathrm{FBP}}|$.

(d) $|\mathfrak{n} - \mathfrak{n}_{\mathrm{TV}-\ell_2}|$.

Figure 9: Fiber reconstruction for MSNR $= 20\,\mathrm{dB}$ and $N_\theta = 90$. Reconstructed image $\widetilde{\mathfrak{n}}$ using (a) FBP and (b) TV-ℓ_2 reconstruction methods. Difference between ground truth $\mathfrak{n}$ and Reconstructed image $\widetilde{\mathfrak{n}}$ using (c) FBP and (d) TV-ℓ_2 reconstruction methods.

threshold value of 10^{-5}, TV-ℓ_2 needs a higher number of iterations, and therefore more time than FBP. However, the reconstruction quality is clearly higher when using our regularized method. In the case where the quality of the image reconstruction is sufficiently high, the threshold can be decreased to a less resctrictive value. At the end of this section we perform a convergence analysis for different threshold values, which allows us to choose the suitable threshold for a required quality or convergence time.

In Fig. 9 we present the reconstructed images for the bundle of fibers for a moderately noisy sensing (MSNR $= 20\,\mathrm{dB}$). We also show the error images in order to provide a better appreciation of the difference between both methods.

In this noisy scenario, we can notice the borders are no longer well estimated using FBP and, as we have coverage of only 25%, some oscillating (Gibbs) artifacts appear. On the contrary, the regularized method provides a good estimation on the borders, with no visible artifacts. We notice certain loss in the dynamics of the image, which causes a lower RSNR compared to the noiseless scenario.

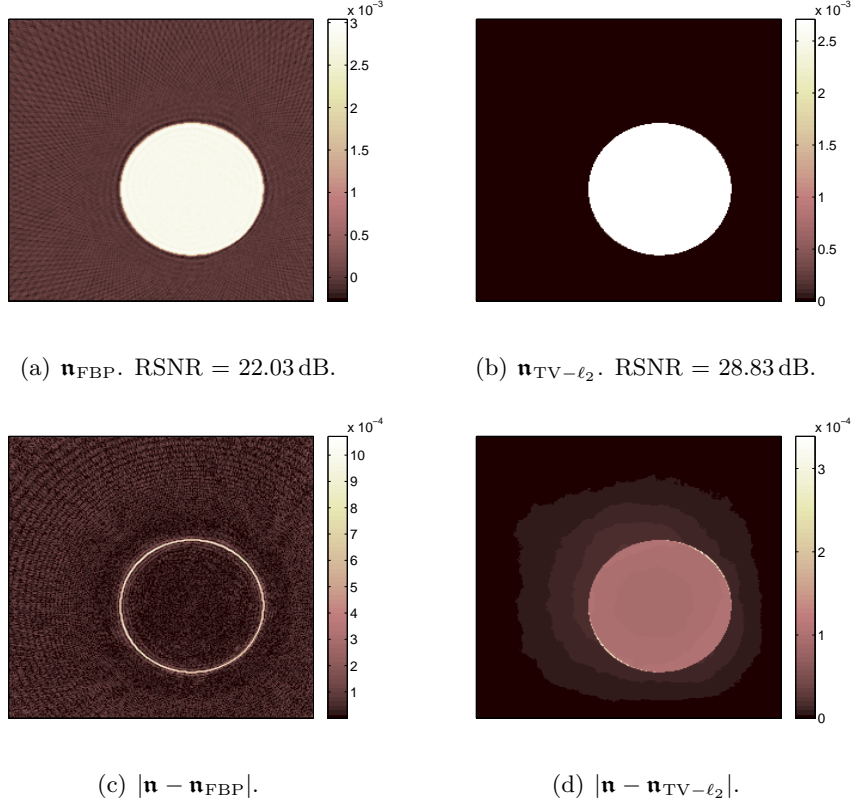Let us show now some results obtained for the other two synthetic images. Fig. 10 and

(a) $\mathfrak{n}_{\text{FBP}}$. RSNR $= 22.03\,\text{dB}$.

(b) $\mathfrak{n}_{\text{TV}-\ell_2}$. RSNR $= 28.83\,\text{dB}$.

(c) $|\mathfrak{n} - \mathfrak{n}_{\text{FBP}}|$.

(d) $|\mathfrak{n} - \mathfrak{n}_{\text{TV}-\ell_2}|$.

Figure 10: Sphere reconstruction for MSNR $= 20\,\text{dB}$ and $N_\theta = 90$. Reconstructed image $\widetilde{\mathfrak{n}}$ using (a) FBP and (b) TV-ℓ_2 reconstruction methods. Difference between ground truth $\mathfrak{n}$ and Reconstructed image $\widetilde{\mathfrak{n}}$ using (c) FBP and (d) TV-ℓ_2 reconstruction methods.

Fig. 11 present the results obtained using FBP and TV-ℓ_2 on the sphere and the Shepp-Logan phantoms, respectively. This comparison was performed for MSNR $= 20\,\text{dB}$ and for $N_\theta = 90$, *i.e.*, a coverage of 25% of the frequency plane.

The regularized method provides a better image dynamics for these phantoms than when reconstructing the fibers for the same noise level. For these phantoms, it can also be observed that FBP reconstructions present a poor estimation on the borders and oscillating artifacts. Moreover, the error image shows a higher discordance with respect to the actual image.

Table 2 presents a more complete comparison of the RSNR obtained using FBP and TV-ℓ_2 on the three synthetic images. The methods are analyzed for three scenarios, one noiseless with MSNR $= \infty$, and the other two with noise such that we have MSNR $= 20\,\text{dB}$ and MSNR $= 10\,\text{dB}$. Results are presented for $\text{Th} = 10^{-5}$ and $N_\theta = 90$, *i.e.*, a coverage of 25% of the frequency plane.

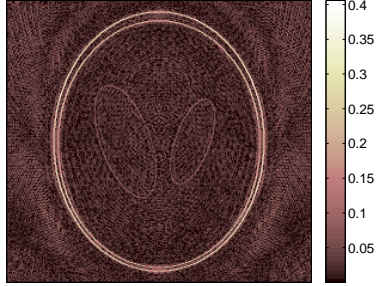| | RSNR[dB] | | | | | |
| | MSNR $= \infty$ | | MSNR $= 20\,\text{dB}$ | | MSNR $= 10\,\text{dB}$ | |
	TV-ℓ_2	FBP	TV-ℓ_2	FBP	TV-ℓ_2	FBP
Fibers	64.54	12.87	26.34	12.77	19.20	11.91
Sphere	34.65	22.34	28.83	22.03	21.85	19.61
Shepp-Logan	39.13	14.19	35.38	14.04	25.33	12.87

Table 2: Comparison of the different RSNR obtained using FBP and TV-ℓ_2 on the three synthetic images for $N_\theta = 90$.
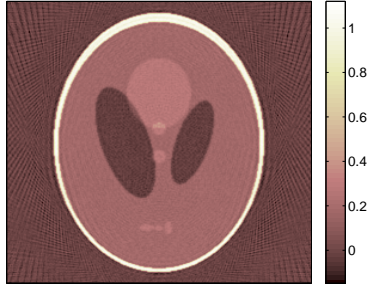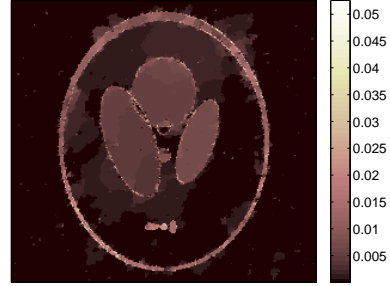
(a) $\mathfrak{n}_{\mathrm{FBP}}$. RSNR $= 14.04\,\mathrm{dB}$.

(b) $\mathfrak{n}_{\mathrm{TV}-\ell_2}$. RSNR $= 35.38\,\mathrm{dB}$.

(c) $|\mathfrak{n} - \mathfrak{n}_{\mathrm{FBP}}|$.

(d) $|\mathfrak{n} - \mathfrak{n}_{\mathrm{TV}-\ell_2}|$.

Figure 11: Shepp-Logan phantom reconstruction for MSNR $= 20\,\mathrm{dB}$ and $N_\theta = 90$. Reconstructed image $\widetilde{\mathfrak{n}}$ using (a) FBP and (b) TV-ℓ_2 reconstruction methods. Difference between ground truth $\mathfrak{n}$ and Reconstructed image $\widetilde{\mathfrak{n}}$ using (c) FBP and (d) TV-ℓ_2 reconstruction methods.
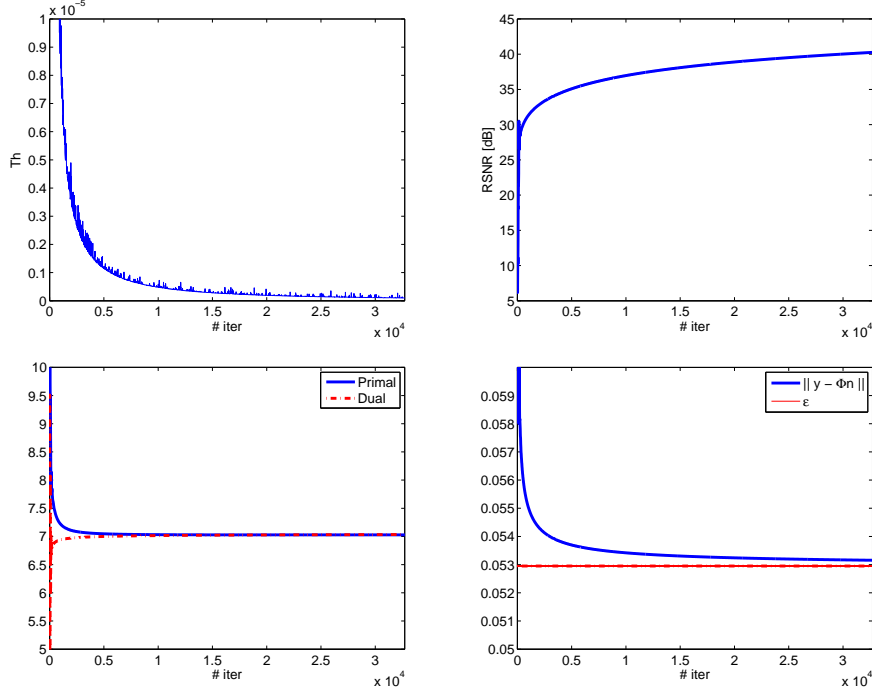
Figure 12: Convergence results when reconstructing the bundle of fibers with $N_\theta = 360$ and MSNR $= 20$ dB. The progress of (top left) the threshold value Th and (top right) the RSNR, along the iterations. (bottom left) The PDG (i) and (bottom right) the condition (ii) of the primal-dual gap corresponding to the indicator function $\iota_\mathcal{C}$.

When comparing the behavior of the algorithm for the different synthetic images, we can notice that TV-ℓ_2 method outperforms the FBP method for all cases. When using the TV-ℓ_2 method it is important to see that, for the noiseless scenario, the fibers are best reconstructed, preserving the image dynamics; while for the noisy scenarios, the Shepp-Logan led to a better image reconstruction.

8.1.3 Algorithm Convergence

Finally, we analyze the convergence of the algorithm by studying the evolution of the primal-dual gap and of the RSNR for different threshold values. For this, we use the reconstruction of the bundle of fibers for $N_\theta = 360$ and MSNR $= 20$ dB.

The evolution of $\|x^{(k+1)} - x^{(k)}\|/\|x^{(k)}\|$ along the iterations is depicted in Fig. 12-(top left) and it helps us to analyze the stability of the algorithm. We can see the curve is not smooth, which indicates a non stable behavior mainly due to a bad conditioning of the global operator K in the product space optimization. This could be improved by a preconditioning procedure as described in [2, 29]. Fig. 12-(top right) presents the RSNR evolution along the iterations.

The evolution of the primal-dual gap along the iterations on the current primal and dual solution is also analyzed by evaluating the partial PDG (pPDG) from (23) and by monitoring the compliance of the conditions required by the indicator functions (24). The behavior of the partial PDG can be observed in Fig. 12-(bottom left). Fig. 12-(bottom right) presents the value of $\|y - \Phi x\|$ compared to the value of ε, which allows us to analyze the compliance of the indicator function $\iota_\mathcal{C}$ in condition (24a). The other conditions are not presented here but we observed that they are satisfied in few iterations.

Table 3 presents the values of some convergence parameters (Time, RSNR, pPDG, ...) for different threshold values. The parameter pPDG % refers to the relative pPDG, computed as pPDG$\% = 100((\text{P} - \text{D})/(\text{P} + \text{D}))$. These results show that a lower threshold provides higher

reconstruction quality but significantly increases the number of CP iterations. In the meantime, the data consistency $\|\boldsymbol{y} - \boldsymbol{\Phi x}\| - \varepsilon$ also decreases. In this specific reconstruction, setting the threshold to $\mathcal{O}(10^{-5})$ or running more than 1000 CP iterations guarantees a RSNR higher than 31 dB.

Th	# iter	Time	RSNR [dB]	pPDG	pPDG [%]	$\|\boldsymbol{y} - \boldsymbol{\Phi x}\| - \varepsilon$
10^{-4}	400	8'	29.52	0.64	4.43	3.7×10^{-3}
10^{-5}	900	18'	31.07	0.31	2.19	2.2×10^{-3}
10^{-6}	5630	2h	35.4	0.04	0.29	6.9×10^{-4}
10^{-7}	32720	11h	40.26	0.005	0.04	2.1×10^{-4}

Table 3: Convergence results when reconstructing the 10 fibers synthetic image with $N_\theta = 360$ and MSNR = 20 dB.

8.1.4 TV-ℓ_2 problem with one constraint

We analyze now the impact of the constraints added to the TV-ℓ_2 formulation in (15). For this, we solve a TV-ℓ_2 problem without the positivity and zero borders constraints, which were proved to guarantee the existence of an unique solution. We call this problem the incomplete TV-ℓ_2 program and we compare it with the complete TV-ℓ_2 program from Sec. 5. For the noiseless scenario, we obtain better reconstruction quality when using the complete version. For the noisy scenarios, the reconstruction results are similar for both versions of the TV-ℓ_2 program. However, the reconstruction for the incomplete program is slower, taking 5 times more iterations to converge than what the complete program takes. We conclude that, in addition to guaranteeing the unicity of the solution, the extra constraints stabilize also the convergence of the reconstruction.

8.2 Experimental Data

The reconstruction algorithm was tested with two particular transparent objects similar to the synthetic data studied in the previous section: a bundle of 10 fibers and a homogeneous sphere, both immersed in an optical fluid. The reconstruction is based on $N_\tau = 696$ parallel and equally spaced light rays. The experimental setup is based on the Schlieren Deflectometric Tomography described in Sec. 2.

A 696×523 pixels CCD camera was used for the acquisition, covering a field of view of 3.25mm $\times$ 2.43mm. This corresponds to $N_\tau = 696$ parallel light rays and 523 2-D slices, which leads to $\delta\tau = 4.7 \times 10^{-3}$mm, and thus to $\delta r = 4.7 \times 10^{-3}$mm.

The experimental configuration leads to a calibration problem. As the object is rotating, the center of rotation may change within a small range. Therefore, a calibration method was implemented as a post-acquisition phase to determine where the projected image is actually centered. The method computes the location of the maximum and minimum deflection values in the direction of τ for all projection angles. Then, the center of the projection image is given by the mean value between these locations.

The characteristics of the objects and the reconstruction results obtained from the measured data are given below.

Bundle of fibers

The first measured object is a bundle of 10 fibers of 200 μm diameter each. The refractive index difference with respect to the solution where the fibers are immersed is $\delta\mathfrak{n} = 12 \times 10^{-3}$.

The experimental data was measured for 60 angular positions over 360 degrees (*i.e.*, 17% of measurements). Fig. 13 shows the reconstruction results obtained using both FBP and TV-ℓ_2 reconstruction methods.
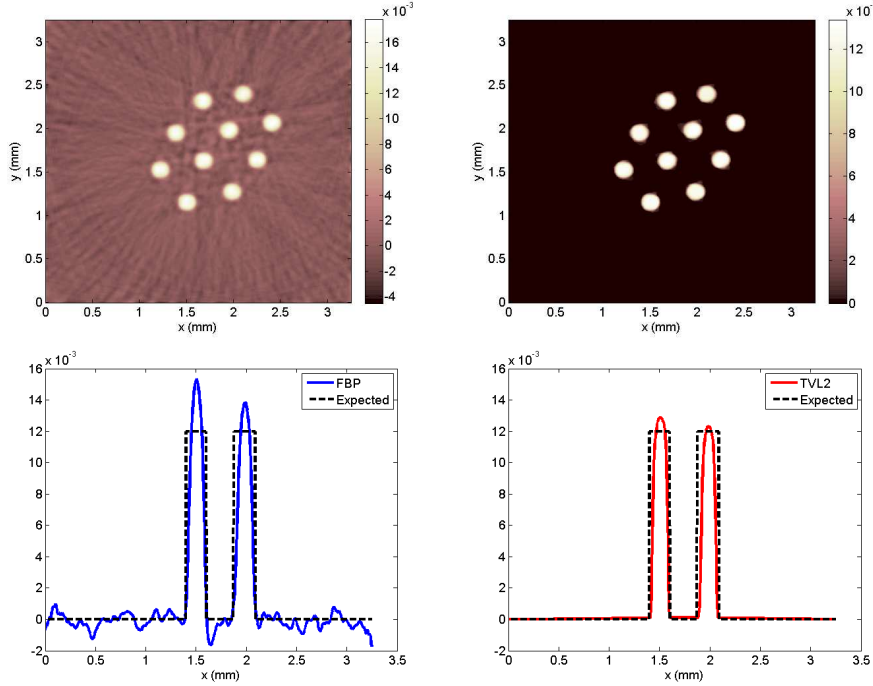

Figure 13: Reconstructed bundle of 10 fibers for 60 angular positions. 2D distribution using (top left) FBP and (top right) TV-ℓ_2. 1D profile along y = 1.218 mm using (bottom left) FBP and (bottom right) TV-ℓ_2.

We observe a similar behavior to the one found in the synthetic reconstruction. We have a more accurate image dynamics recovery using our regularized method, whereas with FBP the reconstruction presents several artifacts and we observe the presence of implausible negative values. It is important to notice that the preservation of the image dynamics depends mainly on the noise estimation and on the proper definition of the constants included in the operator $\boldsymbol{D}$ (see Sec. 3). When considering the appropriate constants, we are able to make an equivalence between the physical problem and its discrete mathematical formulation.

Note that the interior of the reconstructed object is not entirely homogeneous, such reconstruction errors are still present because the modeling error was not taken into consideration. Given a high refractive index difference $\delta\mathfrak{n}$, the first order paraxial approximation is no longer valid and thus the modeling error becomes more important in (1). The actual light ray trajectory could be estimated and inserted in an iterative process as done by Antoine et al. [1]. However, we would then loose the fast computation of the forward model in the frequency domain. Moreover, a section of the bundle of fibers represents a complex image to reconstruct, as the light enters and comes out of multiple fibers causing an error propagation.

As the dimensions of the problem are higher in the experimental setup, the reconstruction time increases. For a threshold of 10^{-6}, FBP converges in 50×10^3 iterations and TV-ℓ_2 in 44×10^3 iterations.

Homogeneous Sphere

The second observed object consists of a homogeneous sphere with a diameter of 1.55 mm. The difference of refractive index between the sphere and the optical fluid where it is immersed is

$\delta\mathfrak{n} = 2.8 \times 10^{-3}$. The deviation map was measured for $N_\theta = 45$ angular positions over 360 degrees (*i.e.*, 13% of measurements). Fig. 14 shows the reconstruction results obtained when using both FBP and TV-ℓ_2 reconstruction methods.
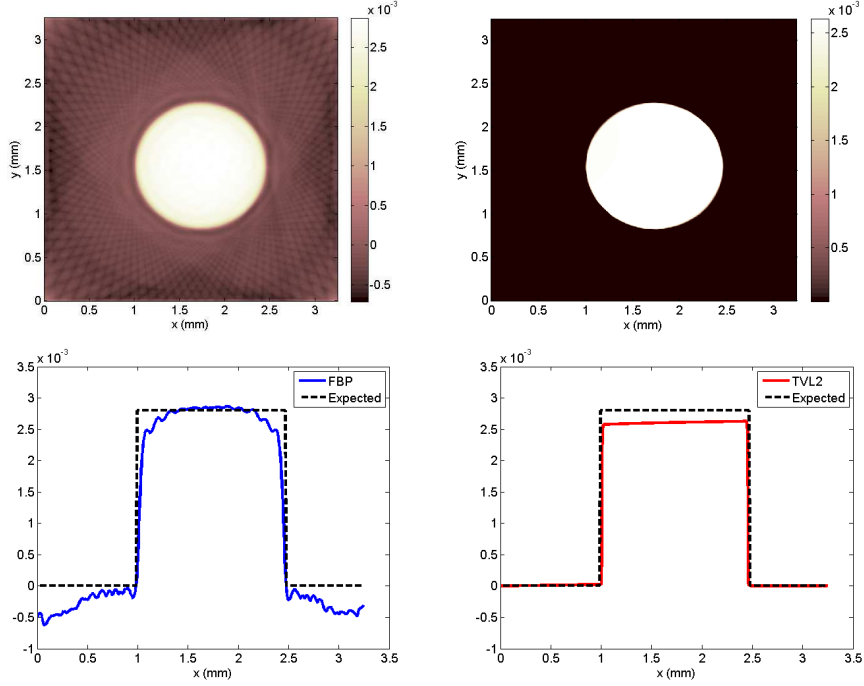

Figure 14: Reconstructed sphere for 45 angular positions. 2D distribution using (top left) FBP and (top right) TV-ℓ_2. 1D profile along y = 1.625 mm using (bottom left) FBP and (bottom right) TV-ℓ_2.

We note similar results as the obtained when reconstructing the fibers: a more accurate image dynamics recovery using our regularized method than FBP. However, when reconstructing the sphere with the TV-ℓ_2 method, we can notice that the interior of the recovered object is more homogeneous and closer to the expected value. This is due to the lower refractive index difference between the sphere and the liquid where it is immerged, which provides a smaller deviation angle and, therefore, a smaller modeling error. Moreover, the sphere represents a less complex image compared to the fibers, as the light enters and comes out of only one object, thus there is no error propagation.

The results are shown for a threshold $\mathtt{Th} = 10^{-6}$, where FBP converges in 50×10^3 iterations and TV-ℓ_2 in 40×10^3 iterations.

9   CONCLUSIONS

We have demonstrated how regularized reconstruction methods, such as TV-ℓ_2, can be used in the framework of Optical Deflectometric Tomography in order to tackle the lack of deflectometric observations and to make the ODT imaging process robust to noise. The proposed constrained optimization problem shows significant improvements in the reconstructions, compared to the well-known Filtered Back Projection. The results confirm that when dealing with a compressive setting, the Total Variation regularization and the prior information constraints (positivity and FoV restriction) help in providing a unique and accurate estimation of the RIM, promoting sharp edges and preserving the image dynamics. By working with the Chambolle-Pock algorithm we exploit the advantages of proximal operators and of primal-dual algorithms, and their flexibility to integrate multiple constraints. We have also shown that the use of the fast NFFT algorithm

30

efficiently approximates (with a controlled error) the ODT sensing model involving the polar NDFT.

Noticeably, there still exist some artifacts in the experimental data reconstruction, coming from the modeling error. In order to handle this problem, we should no longer assume a linear light propagation. The actual curved light trajectories, depending themselves on the RIM through the eikonal equation, must be traced. However, such a model improvement breaks the fast computability of the forward sensing operator through the Fourier domain and knowing how to insert this new scheme in our reconstruction method is a matter of future works.

We have also noticed the long convergence time of the algorithm. As the CP algorithm has been proved to work slowly when the problem is badly conditioned [2, 29], convergence results could be improved by preconditioning our global operator $\boldsymbol{K}$. We note also that most of the algorithms are implemented in Matlab, except for the NFFT algorithm which is compiled in C++. This makes the implementation slower, and could be improved for instance by the use of parallel processing techniques.

We finally remark that the Optical Deflectometric framework treated in this paper can be also applied to other imaging techniques, such as the X-ray phase contrast tomography [28]. With the use of the energetic X-ray light, the first order paraxial approximation involving linear light trajectory is no longer needed and the proposed methods are expected to provide better results.

ACKNOWLEDGEMENTS

A DEFLECTOMETRIC FOURIER SLICE THEOREM

Using the notations of Sec. 2.1, we must prove that

$$y(\omega, \theta) \;=\; \tfrac{2\pi i \omega}{\mathrm{n_r}}\, \widehat{\mathfrak{n}}\big(\omega\, \boldsymbol{p}_\theta\big).$$

By definition of y in (3), we have

$$y(\omega, \theta) = \tfrac{1}{\mathrm{n_r}} \int_{\mathbb{R}} \int_{\mathbb{R}^2} \big(\boldsymbol{\nabla}\mathfrak{n}(\boldsymbol{r}) \cdot \boldsymbol{p}_\theta\big)\, \delta(\tau - \boldsymbol{r} \cdot \boldsymbol{p}_\theta)\, e^{-2\pi i\, \tau \omega}\, \mathrm{d}^2\boldsymbol{r}\, \mathrm{d}\tau$$

$$= \tfrac{1}{\mathrm{n_r}} \int_{\mathbb{R}^2} \big(\boldsymbol{\nabla}\mathfrak{n}(\boldsymbol{r}) \cdot \boldsymbol{p}_\theta\big)\, e^{-2\pi i\, \boldsymbol{r} \cdot (\omega \boldsymbol{p}_\theta)}\, \mathrm{d}^2\boldsymbol{r}.$$

However, for any function $f : \mathbb{R} \to \mathbb{C}$ integrable on $\mathbb{R}$,

$$\widehat{\tfrac{\mathrm{d}f}{\mathrm{d}t}}(\omega) = \int_{\mathbb{R}} \tfrac{\mathrm{d}}{\mathrm{d}t} f(t)\, e^{-2\pi i\, t \omega}\, \mathrm{d}t = (2\pi i)\, \omega\, \hat{f}(\omega).$$

Therefore, we compute easily for any $\boldsymbol{a}, \boldsymbol{\xi} \in \mathbb{R}^2$,

$$\int_{\mathbb{R}^2} \big(\boldsymbol{\nabla}\mathfrak{n}(\boldsymbol{r}) \cdot \boldsymbol{a}\big)\, e^{-2\pi i\, \boldsymbol{r} \cdot \boldsymbol{\xi}}\, \mathrm{d}^2\boldsymbol{r} = (2\pi i)\, (\boldsymbol{\xi} \cdot \boldsymbol{a})\, \hat{\mathfrak{n}}(\boldsymbol{\xi}).$$

Setting $\boldsymbol{a} = \boldsymbol{p}_\theta$ and $\boldsymbol{\xi} = \omega \boldsymbol{p}_\theta$, we find finally

$$y(\omega, \theta) = \tfrac{1}{\mathrm{n_r}} \int_{\mathbb{R}^2} \big(\boldsymbol{\nabla}\mathfrak{n}(\boldsymbol{r}) \cdot \boldsymbol{p}_\theta\big)\, e^{-2\pi i\, \boldsymbol{r} \cdot (\omega \boldsymbol{p}_\theta)}\, \mathrm{d}^2\boldsymbol{r} = \tfrac{2\pi i \omega}{\mathrm{n_r}}\, \hat{n}(\omega \boldsymbol{p}_\theta).$$

B NON-EQUISPACED FOURIER TRANSFORM

The non-equispaced Fourier Transform (NFFT) allows us to compute rapidly, *i.e.*, in $\mathcal{O}(N \log N)$, the NDFT defined in (8) of a function defined on $\mathcal{C}_N$. This computation is performed with a controllable error which can be further reduced by increasing the computational time.

In a nutshell, the NFFT algorithm replaces the equivalent matrix multiplication $\hat{\boldsymbol{f}} = \boldsymbol{F}\boldsymbol{f}$ of (9) by $\hat{\boldsymbol{s}} = \widetilde{\boldsymbol{F}}\boldsymbol{f}$, where $\widetilde{\boldsymbol{F}}$ has fast matrix-vector multiplication computation. In this scheme, the discrepancy between $\hat{\boldsymbol{f}}$ and $\hat{\boldsymbol{s}}$, as measured by $E_\infty(\boldsymbol{f}) := \|\hat{\boldsymbol{f}} - \hat{\boldsymbol{s}}\|_\infty$, is controlled and kept small.

More precisely, the matrix $\widetilde{\boldsymbol{F}}$ starts by embedding $\mathbb{R}^N$ in a bigger regular space $\mathbb{R}^n$ of size $n = n_0^2$, with $n_0 > N_0$. This is obtained from the multiplication of $\boldsymbol{f}$ with a matrix $\boldsymbol{W} \in \mathbb{R}^{n \times N}$ that performs a weighting of $\boldsymbol{f}$ by a vector $\boldsymbol{w} \in \mathbb{R}_+^N$ (component wise) followed by a symmetric zero-padding on each side of the function domain $\mathcal{C}_N$. Once in $\mathbb{R}^n$, the common DFT matrix $\boldsymbol{F}_n$ is applied. It can be computed with the FFT algorithm in order to obtained an oversampled Fourier transform of $\boldsymbol{f} \in \mathbb{R}^N$. Finally, a sparse matrix $\boldsymbol{V} \in \mathbb{R}^{M \times n}$ multiplies the output of $\boldsymbol{F}_n$ in order to end in the M dimensional space $\hat{\mathcal{P}}$. Each row of $\boldsymbol{V}$ corresponds to the translation in the 2-D Fourier domain of a compact and separable 2-D filter ψ on one specific point of the non-regular grid $\hat{\mathcal{P}}_M$. As a final result, the matrix $\widetilde{\boldsymbol{F}}$ is thus factorized in [24]

$$\widetilde{\boldsymbol{F}} = \boldsymbol{V}\boldsymbol{F}_n\boldsymbol{W}.$$

Without entering into unnecessary technicalities, the NFFT scheme is characterized by a precise connection between the function ψ defining $\boldsymbol{V}$ and the weighting performed by $\boldsymbol{W}$. In particular, each component of $\boldsymbol{w}$ is actually set to the inverse of the Fourier transform of a filter φ, while ψ is a periodization of (a truncation of) the same filter.

There exist several choices of windows φ/ψ associated to different numerical properties (*e.g.*, localized support in frequency and time, simple precomputations of the windows, ...). We select here the *translates of Gaussian bells* [33], involving a Gaussian behavior for $\hat{\varphi}(k)$, which provides fast error decay for $E_\infty(\boldsymbol{f})$.

In particular, denoting by $\alpha = n/N \geqslant 1$ the oversampling factor and using the FFT for matrix-vector multiplications involving $\boldsymbol{F}_n$, the total complexity $\mathcal{T}$ of the multiplications of $\widetilde{\boldsymbol{F}}$ or $\widetilde{\boldsymbol{F}}^*$ with vectors is

$$\mathcal{T}(\widetilde{\boldsymbol{F}}) = \mathcal{O}(n \log n + N(\tfrac{2\alpha - 1}{2\alpha - 2}) \log 1/\epsilon),$$

if we impose $E_\infty(\boldsymbol{f}) \leqslant 4\|\boldsymbol{f}\|_1\,\epsilon$. For a fixed $\alpha > 1$, this reduces to $\mathcal{T}(\widetilde{\boldsymbol{F}}) = \mathcal{O}(N \log N/\epsilon)$, which is far less than a direct computation of the DFT in $O(MN)$ computation even for small value of ϵ.

C CONVEX CONJUGATE FUNCTIONS

C.1 Convex conjugate of the function G.

Given a vector $\boldsymbol{t} = (\boldsymbol{t}_1^T, \boldsymbol{t}_2^T)^T \in \mathbb{R}^{2N}$, we can define the function $G(\boldsymbol{t}) = H(\boldsymbol{t}) + \imath_{\Pi_{1,2}}(\boldsymbol{t}_1, \boldsymbol{t}_2)$, with the bisector plane $\Pi_{1,2} = \{\boldsymbol{t} : \boldsymbol{t}_1 = \boldsymbol{t}_2\}$. Therefore, for a vector $\boldsymbol{u} = (\boldsymbol{u}_1^T, \boldsymbol{u}_2^T)^T \in \mathbb{R}^{2N}$, the dual function $G^*(\boldsymbol{t})$ can be computed as follows:

$$\begin{aligned}
G^*(\boldsymbol{t}) &= \max_{\boldsymbol{u}} \langle \boldsymbol{u}, \boldsymbol{t} \rangle - G(\boldsymbol{t}) \\
&= \max_{\boldsymbol{u}:\ \boldsymbol{u}_1 = \boldsymbol{u}_2} \langle \boldsymbol{u}_1, \boldsymbol{t}_1 + \boldsymbol{t}_2 \rangle - H(\boldsymbol{u}_1) \\
&= \begin{pmatrix} \mathbb{I}^N \\ \mathbb{I}^N \end{pmatrix} H^*(\boldsymbol{t}_1 + \boldsymbol{t}_2).
\end{aligned}$$

C.2 Convex Conjugate of the indicator function $\imath_{\mathcal{C}}$.

Given the indicator function $\imath_{\mathcal{C}}(\boldsymbol{u})$ of the convex set $\mathcal{C} = \{\boldsymbol{v} \in \mathbb{R}^M : \|\boldsymbol{v} - \boldsymbol{y}\| \leqslant \varepsilon\}$, its dual function can be computed via the Legendre transform as follows:

$$
\begin{aligned}
\imath_{\mathcal{C}}^*(\boldsymbol{v}) &= \max_{\boldsymbol{u}} \ \langle \boldsymbol{v}, \boldsymbol{u} \rangle - \imath_{\mathcal{C}}(\boldsymbol{u}) \\
&= \langle \boldsymbol{v}, \boldsymbol{y} \rangle + \max_{\boldsymbol{u}:\|\boldsymbol{u}-\boldsymbol{y}\|\leqslant\varepsilon} \ \langle \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{y} \rangle \\
&= \langle \boldsymbol{v}, \boldsymbol{y} \rangle + \max_{\boldsymbol{b}:\|\boldsymbol{b}\|\leqslant\varepsilon} \ \langle \boldsymbol{v}, \boldsymbol{b} \rangle.
\end{aligned}
$$

The value of $\{\boldsymbol{b} : \|\boldsymbol{b}\| \leqslant \varepsilon\}$ that maximizes the last expression is $\boldsymbol{b} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}\varepsilon$, and we get $\imath_{\mathcal{C}}^*(\boldsymbol{v}) = \langle \boldsymbol{v}, \boldsymbol{y} \rangle + \varepsilon\|\boldsymbol{v}\|$.

D   DETAILS ON THE PRODUCT SPACE OPTIMIZATION

D.1   Proximal operator of the function G

For every $\boldsymbol{t} = (\boldsymbol{t}_1, \cdots, \boldsymbol{t}_p) \in \mathbb{R}^{pN}$, the function $G(\boldsymbol{t})$ is defined as:

$$
G(\boldsymbol{t}) = \sum_{j=2}^{p} \imath_{\Pi_{1,j}}(\boldsymbol{t}) + H(\boldsymbol{t}_1),
$$

where, for $j \in [p]$, $\Pi_{1,j}$ denotes the bisector plane $\Pi_{1,j} = \{\boldsymbol{t} \in \mathbb{R}^{pN} : \boldsymbol{t}_1 = \boldsymbol{t}_j\}$.

Then, for $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \cdots, \boldsymbol{\zeta}_p) \in \mathbb{R}^{pN}$, the proximal operator of G reduces to

$$
\widetilde{\boldsymbol{t}} = \mathrm{prox}_{\mu G} \boldsymbol{\zeta} = \underset{\boldsymbol{t}:\ \boldsymbol{t}_1 = \cdots = \boldsymbol{t}_p}{\arg\min} \ \mu H(\boldsymbol{t}_1) + \tfrac{1}{2}\|\boldsymbol{\zeta} - \boldsymbol{t}\|^2.
$$

As all the $\boldsymbol{t}_j$ are equal, we necessarily have $\widetilde{\boldsymbol{t}} = (\mathbb{I}^N, \cdots, \mathbb{I}^N)^T \widetilde{\boldsymbol{u}}$ with $\widetilde{\boldsymbol{u}} \in \mathbb{R}^N$ defined by

$$
\begin{aligned}
\widetilde{\boldsymbol{u}} &= \underset{\boldsymbol{u}\in\mathbb{R}^N}{\arg\min} \, \mu H(\boldsymbol{u}) + \tfrac{1}{2}\sum_j \|\boldsymbol{\zeta}_j - \boldsymbol{u}\|^2 \\
&= \underset{\boldsymbol{u}\in\mathbb{R}^N}{\arg\min} \, \mu H(\boldsymbol{u}) + \tfrac{1}{2}\big[\|\boldsymbol{\zeta}_1\|^2 + \cdots + \|\boldsymbol{\zeta}_p\|^2 - 2p\,\bar{\boldsymbol{\zeta}}^T\boldsymbol{u} + p\|\boldsymbol{u}\|^2\big] \\
&= \underset{\boldsymbol{u}\in\mathbb{R}^N}{\arg\min} \, \mu H(\boldsymbol{u}) + \tfrac{1}{2}\big[p\|\bar{\boldsymbol{\zeta}}\|^2 - 2p\,\bar{\boldsymbol{\zeta}}^T\boldsymbol{u} + p\|\boldsymbol{u}\|^2\big] \\
&= \mathrm{prox}_{(\mu/p)H}\,\bar{\boldsymbol{\zeta}}.
\end{aligned}
$$

with $\bar{\boldsymbol{\zeta}} = \frac{1}{p}\sum_j \boldsymbol{\zeta}_j \in \mathbb{R}^N$ and where we used the fact that we can always subtract or add constants in the minimization without disturbing its solution. Denoting by $\mathbb{I}_p^N = (\mathbb{I}^N, \cdots, \mathbb{I}^N) \in \mathbb{R}^{N \times pN}$ the operator such that $\mathbb{I}_p^N \boldsymbol{\zeta} = p\bar{\boldsymbol{\zeta}}$, this provides finally the compact notation

$$
\mathrm{prox}_{\mu G}\,\boldsymbol{\zeta} = (\mathbb{I}_p^N)^T \, \mathrm{prox}_{\frac{\mu}{p}H} \, \tfrac{1}{p}\mathbb{I}_p^N \boldsymbol{\zeta}.
$$

D.2 Formulation of Chambolle-Pock algorithm in OD

We take the CP algorithm as described in Eq. (18) and we translate it into our OD problem in the product space, to obtain the following:

$$
\begin{cases}
\boldsymbol{s}^{(k+1)} = \begin{pmatrix} \boldsymbol{s}_1^{(k+1)} \\ \boldsymbol{s}_2^{(k+1)} \end{pmatrix} = \begin{pmatrix} \text{prox}_{\nu F_1^*}\left(\boldsymbol{s}_1^{(k)} + \nu\boldsymbol{\nabla}\bar{\boldsymbol{t}}_1^{(k)}\right) \\ \text{prox}_{\nu F_2^*}\left(\boldsymbol{s}_2^{(k)} + \nu\boldsymbol{\Phi}\bar{\boldsymbol{t}}_2^{(k)}\right) \end{pmatrix}, \\
\boldsymbol{t}^{(k+1)} = \begin{pmatrix} \boldsymbol{t}_1^{(k+1)} \\ \boldsymbol{t}_2^{(k+1)} \end{pmatrix} \\
\qquad\quad = \begin{pmatrix} \mathbb{I}^N \\ \mathbb{I}^N \end{pmatrix} \text{prox}_{\frac{\mu}{2}H} \tfrac{1}{2}(\boldsymbol{t}_1^{(k)} - \mu\boldsymbol{\nabla}^*\boldsymbol{s}_1^{(k+1)} + \boldsymbol{t}_2^{(k)} - \mu\boldsymbol{A}^*\boldsymbol{s}_2^{(k+1)}), \\
\bar{\boldsymbol{t}}^{(k+1)} = \begin{pmatrix} \bar{\boldsymbol{t}}_1^{(k+1)} \\ \bar{\boldsymbol{t}}_2^{(k+1)} \end{pmatrix} = 2\begin{pmatrix} \boldsymbol{t}_1^{(k+1)} \\ \boldsymbol{t}_2^{(k+1)} \end{pmatrix} - \begin{pmatrix} \boldsymbol{t}_1^{(k)} \\ \boldsymbol{t}_2^{(k)} \end{pmatrix}.
\end{cases}
$$

We have the function $H(\boldsymbol{t}) = \imath_{\mathcal{P}_0}(\boldsymbol{t})$ and its proximal operator $\text{prox}_{\frac{\mu}{2}H}\boldsymbol{\zeta} = \text{proj}_{\mathcal{P}_0}\boldsymbol{\zeta}$. As $\boldsymbol{t}_1 = \boldsymbol{t}_2$, we can relabel the variable as $\boldsymbol{x}^{(k)} = \boldsymbol{t}_1^{(k)} = \boldsymbol{t}_2^{(k)}$ and $\boldsymbol{x}^{(k+1)} = \boldsymbol{t}_1^{(k+1)} = \boldsymbol{t}_2^{(k+1)}$. In the same way, $\bar{\boldsymbol{t}}_1 = \bar{\boldsymbol{t}}_2$, thus we can relabel the variable as $\bar{\boldsymbol{x}}^{(k)} = \bar{\boldsymbol{t}}_1^{(k)} = \bar{\boldsymbol{t}}_2^{(k)}$ and $\bar{\boldsymbol{x}}^{(k+1)} = \bar{\boldsymbol{t}}_1^{(k+1)} = \bar{\boldsymbol{t}}_2^{(k+1)}$. We obtain the following algorithm:

$$
\begin{cases}
\boldsymbol{s}_1^{(k+1)} &= \text{prox}_{\nu F_1^*}(\boldsymbol{s}_1^{(k)} + \nu\boldsymbol{\nabla}\bar{\boldsymbol{x}}^{(k)}), \\
\boldsymbol{s}_2^{(k+1)} &= \text{prox}_{\nu F_2^*}(\boldsymbol{s}_2^{(k)} + \nu\boldsymbol{\Phi}\bar{\boldsymbol{x}}^{(k)}), \\
\boldsymbol{x}^{(k+1)} &= \text{proj}_{\mathcal{P}_0}\left(\boldsymbol{x}^{(k)} - \tfrac{\mu}{2}(\boldsymbol{\nabla}^*\boldsymbol{s}_1^{(k+1)} + \boldsymbol{\Phi}^*\boldsymbol{s}_2^{(k+1)})\right), \\
\bar{\boldsymbol{x}}^{(k+1)} &= 2\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}.
\end{cases}
$$

References

[1] P. Antoine, E. Foumouo, J.-L. Dewandel, D. Beghuin, A. González, and L. Jacques, *Optical deflection tomography with reconstruction based on sparsity*, Proc. of OPTIMESS2012, (Antwerp, Belgium), April 2012.

[2] S. Becker and J. Fadili, *A quasi-newton proximal splitting method*, preprint, arxiv:1206.1156, (2012).

[3] D. Beghuin, J.L. Dewandel, L. Joannes, E. Foumouo, and P. Antoine, *Optical deflection tomography with the phase-shifting schlieren*, Optics Letters, **35** (2010), no. 22, 3745–3747.

[4] M. Born, E. Wolf, and A.B. Bhatia, *"Principles of optics: electromagnetic theory of propagation, interference and diffraction of light"*, Cambridge Univ Press, 1999.

[5] S. Boyd and L. Vandenberghe, *"Convex Optimization"*, Cambridge Univ Press, 2004.

[6] E.J. Candes and T. Tao, *Near-optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Transactions on Information Theory, **52** (2006), no. 12, 5406–5425.

[7] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, **20** (2004), no. 1, 89–97.

[8] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, **40** (2011), no. 1, 120–145.

[9] P.L. Combettes and J.C. Pesquet, *Proximal splitting methods in signal processing*, Fixed-Point Algorithms for Inverse Problems in Science and Engineering, (2011), 185–212.

[10] W. Cong, A. Momose, and G. Wang, *Fourier transform-based iterative method for differential phase-contrast computed tomography*, Optics Letters, **37** (2012), no. 11, 1784–1786.

[11] W. Cong and G. Wang, *Differential phase-contrast interior tomography*, preprint, arxiv:1108.1414, (2011).

[12] D.L. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory, **52** (2006), no. 4, 1289–1306.

[13] D.L. Donoho and J.M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, **81** (1994), no. 3, 425–455.

[14] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, T. Sun, K.F. Kelly, and R.G. Baraniuk, *Single-pixel imaging via compressive sampling*, IEEE Signal Processing Magazine, **25** (2008), no. 2, 83–91.

[15] F.X. Dupé, M.J. Fadili, and J.-L. Starck, *Inverse problems with poisson noise: Primal and primal-dual splitting*, IEEE International Conference on Image Processing (ICIP) (Brussels, Belgium), September 2011.

[16] G. Faris and R. Byer, *Beam-deflection optical tomography*, Optics Letters, **12** (1987), no. 2, 72–74.

[17] E. Foumouo, J.L. Dewandel, L. Joannes, D. Beghuin, L. Jacques, and P. Antoine, *Optical tomography based on phase-shifting schlieren deflectometry*, Proceedings of SPIE, **7726** (2010).

[18] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association, **58** (1963), no. 301, 13–30.

[19] L. Jacques, A. González, E. Foumouo, and P. Antoine, *Refractive index map reconstruction in optical deflectometry using total-variation regularization*, Proceedings of SPIE, **8138** (2011).

[20] L. Jacques, D. Hammond, and J. Fadili, *Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine*, IEEE Transactions on Information Theory, **57** (2011), no. 1, 559–571.

[21] L. Joannes, F. Dubois, and J.C. Legros, *Phase-shifting schlieren: high-resolution quantitative schlieren that uses the phase-shifting technique principle*, Applied Optics, **42** (2003), no. 25, 5046–5053.

[22] O. Kafri, *Noncoherent method for mapping phase objects*, Optics Letters, **5** (1980), no. 12, 555–557.

[23] A. C. Kak and Slaney M., *"Principles of computerized tomographic imaging"*, IEEE Press, 1988.

[24] J. Keiner, S. Kunis, and D. Potts, *Using NFFT 3—a software library for various nonequispaced fast fourier transforms*, ACM Transactions on Mathematical Software (TOMS), **36** (2009), no. 4.

[25] M. Lustig, D. Donoho, and J.M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnetic Resonance in Medicine, **58** (2007), no. 6, 1182–1195.

[26] J.J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien. (French) [Dual convex functions and proximal points in a Hilbert space]*, CR Acad. Sci. Paris, **255** (1962), 2897–2899.

[27] K. Morita, *"Applied Fourier Transform"*, IOS Press, 1995.

[28] F. Pfeiffer, C. Kottler, O. Bunk, and C. David, *Hard X-ray phase tomography with low-brilliance sources*, Physical Review Letters, **98** (2007), no. 10.

[29] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, ICCV, (2011), 1762–1769.

[30] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, *Improved total variation-based CT image reconstruction applied to clinical data*, Physics in Medicine and Biology, **56** (2011), 1545–1561.

[31] L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, **60** (1992), no. 1-4, 259–268.

[32] E.Y. Sidky, J.H. Jørgensen, and X. Pan, *Convex optimization problem prototyping with the Chambolle-Pock algorithm for image reconstruction in computed tomography*, preprint, arxiv:1111.5632, (2011).

[33] G. Steidl, *A note on fast fourier transforms for nonequispaced grids*, Advances in Computational Mathematics, **9** (1998), no. 3, 337–352.

[34] C. Taswell, *The what, how, and why of wavelet shrinkage denoising*, Computing in Science & Engineering, **2** (2000), no. 3, 12–19.

[35] Y. Wiaux, L. Jacques, G. Puy, AMM Scaife, and P. Vandergheynst, *Compressed sensing imaging techniques for radio interferometry*, Monthly Notices of the Royal Astronomical Society, **395** (2009), no. 3, 1733–1742.

[36] C. Zheng, J. Wu, and W. Sun, *Applications of pseudo-polar FFT in synthetic aperture radar imaging*, PIERS Online, **3** (2007), no. 1, 25–30.